

Recipe for ex-Soviet republics' science

The republics that were formerly part of the Soviet Union embark on a journey full of hazard: sooner or later they will have to transform the structure of science left by Mr Mikhail Gorbachev.

WHETHER Mr Mikhail Gorbachev, who left the Kremlin last week, was a greater success than a failure is a version of the old difficulty of comparing very large numbers: both his successes and his failures were huge. His use of the absolute power he enjoyed six years ago to dissolve the Soviet empire in Central Europe, his doctrine of *glasnost* (which has at least allowed his downfall last week to be fully chronicled) and his willingness to negotiate deals on arms control in Europe and more generally, driven though that may have been by the ruinous cost of sustaining global strategic power in the shambles of the old Soviet economy, are memorable. (The course of recent history, and even Gorbachev's career, might have been even more cheerful if President Ronald Reagan's advisers had let him hold to the plan for getting rid of all nuclear weapons to which he had briefly agreed at Reykjavik in 1987.) Gorbachev's outstanding claim on public attention is that he brought an unprecedented degree of personal intelligence to public life, and gave it almost free rein. That in the process he changed the world is not disputed.

Gorbachev's failures stem from two conspicuous errors, of which the chief was his failure to distance himself from the Communist Party of the old Soviet Union. Although his first power base was that of general secretary of the Party, by 1987 he was also president of the union; it is mystifying that such a clever person never sensed the conflict of interest between that national office and the supposition that the Party should have a constitutionally 'leading role'. Yet the regular bruising meetings of the Politburo and the Central Committee were the means by which brave plans were emasculated by tawdry compromise. (The enmity between Gorbachev and Boris Yeltsin, now the president of Russia, dates from such a meeting.) Gorbachev's other serious error has been his failure, over seven years, to grasp the nettle of economic reform. His consistent argument was that legal and constitutional reform should come first, but that has proved to be a recipe for bankruptcy.

Bankruptcy

Bankruptcy is the most perplexing of the legacies Gorbachev's successors will inherit. In Russia, the reform process is due to begin today (2 January) with the unfreezing of all prices in the retail trade and something like a fivefold increase in the cost of fuel. Unless this change unlocks previously unknown hoards of food and other

supplies, it will exacerbate an already dangerous situation. Market prices are those that strike a balance between supply and demand, but if the supply is physically insufficient to keep the people alive through the winter now begun in earnest, those on the edge of starvation will not be comforted that their plight will send a strong economic signal to those responsible for next year's harvest. Yeltsin and his fellow-presidents will need more than mere nerve; they will need a convincing demonstration that they will be in better shape during the winters ahead, if they are to command the fortitude of their electors in the hard months ahead.

Ignorance

How far-sighted will they be? One danger is that the nationalist scramble for the assets of the old union (Ilyushin aircraft, for example) will obscure simpler economic truths (that capital assets are of literally no value if they cannot be used, for example). Another, more serious, is that the republics are bent on embracing what they call a market economy without a proper sense of what is entailed. After 75 years, understanding of the interplay between capital and labour seems to have been thoroughly obliterated. Despite *Das Kapital* and all that flowed from it, the notion that capital costs money remains mystifying (and, when understood, is as often resented).

The civilities indispensable to a stable market economy are non-existent. There is, for example, no way of collecting income tax except by deductions from the salaries of those who happen to be on government payrolls, the new sales tax (at a rate of 28 per cent) will test the honesty of Russia's retailers to the full — and there is no system of social security for relieving the hardship of those thrown out of work by economic change. The republics now rushing towards reform may be likened to a group of people embarking by light aircraft on a journey across an uncharted continent without navigational aids or even an understanding of aerodynamics. Let us hope they enjoy the luck that they will need.

The hazards of such a journey may require that many valuable things will have to be jettisoned on the way. The international scientific community should have a particular regard that an early casualty may be the distinctive Soviet scientific enterprise. The prospect seems to be (see *Nature* 354, 499; 19 December 1991) that the structure of the Soviet Academy of Sciences will kept going, perhaps



hand to mouth, as the Russian academy. As a stopgap, that is probably necessary, but outside help will be essential if able people are to remain productive (or even where they are). But the structure will not suffice for the long run if the Russian government endorses even a small part of the old Marxist rhetoric about the importance of science in the modern state. Radical upheaval will be required.

Here is why, and how. The present system in which a network of research institutes (now truncated by the departure of the Ukraine) is supported and managed directly by an academy embodies an intolerable conflict of interest. Especially in such rough times, the interests of the academy are indistinguishable from those of the directors of its institutes, traditionally personal fiefdoms in which the power of personal patronage, even if moderated by the notion that the institutes will be collectives, will now be greater than ever. These are hardly the circumstances in which basic research can flourish, nor will they help to create a research enterprise of the kind that a would-be modern state such as Russia needs. The academy's complacent assumption that only its own institutes are worthwhile is at once mistaken (many of its institutes are second-rate) and a recipe for the indulgence of self-interest. (The old academy's attempts to set up a system of competitive research grants might have been more successful if its own payroll pensioners had been less influential in its councils.) When this transitional year is over, it should withdraw from the direct management of research.

There is plenty of other work to be done. What, for example, is to happen to the universities which, despite the efforts of Gorbachev's able minister Yagodin, are still largely cast in their Stalinist mould, itself a means of keeping students away from knowledge judged irrelevant to their immediate needs as well as from the traditional subversiveness of teachers? An academy with the national interest (Russia's now) at heart could have a decisive influence on the overdue restructuring of the universities — and on the encouragement of research therein.

There is a similar need in the new republics' reconstruction of the apparatus of industrial research. During 75 years of central planning, production ministries have maintained research institutes to serve the needs of their own production plants, while the Soviet academy's institutes were regarded as a source of radical innovations that might, after negotiation within a labyrinthine bureaucracy, be scheduled for production at a factory whose managers might not previously have known what was in the wind.

There could hardly have been a mechanism for what is called technology transfer less suitable for matching the performance of new products to market needs. The notion that academy institutes, functioning as collectives, may now make deals with production enterprises may help a little, but that random process will do little to equip Russia or any of the other republics with the skills in research and development they will eventually need. What, in the meantime, can be done to keep alive and to improve a hard

core of industrial research and development? The Russian academy could more easily bend its mind to that important issue if it were freed from the anxiety of keeping its own institutes in being.

To blame Gorbachev for not having resolved these problems while still in office would be unfair. He had other things on his mind. But it is a fact that nothing much was done in his time to change a structure of research so cumbersome and unsuited to its declared purposes that it is almost miraculous that so much talent has survived in spite of it. With all the social upheavals now in prospect, the decades ahead will not be so lucky unless reform comes soon.

Opening for US science

US scientific organizations want (and need) to find ways to help colleagues in the former Soviet Union.

THE crisis in the republics of what was the Soviet Union presents a special challenge to scientists in Western nations who want to see science not only survive but also improve as an enterprise during the coming years. The new Russian Republic, home to the majority of the most able Soviet scientists, is of special concern, particularly in the light of reports that as many as half of Russia's most cited scientists have left the country. What are the immediate needs? What can be done?

Although the organizational structure of Russian science is in disarray (former Soviet colleagues are no longer in positions of authority or lack resources), various groups of US scientists are trying to maintain contact. Recently, the US National Academy of Sciences (NAS) sent a delegation from its committee on dual-use technologies to talk about future industries. Concern about the fact that Russian scientists (like those in the Ukraine and elsewhere) have no currency to purchase essential reagents, spare parts and the like, has led to conversations between the NAS and the US government, as well as the Royal Society in Britain, about ways to help. But no decisions have been reached.

On 10 January, a group representing wealthy US foundations, including the Rockefeller, Carnegie, MacArthur and Ford Foundations, will leave for Moscow on an exploratory mission. With the ruble virtually worthless, contributions of even a few millions of dollars could have a major impact on the professional lives of Soviet scientists.

The American Academy of Arts and Sciences, whose headquarters are in Boston, is also thinking about ways to help. One plan under consideration is to form a clearing house for information on grants and potential collaborations. Provisions in certain US research grants, for instance, permit as much as 5 per cent of funds to be allocated to work with scientists from other countries.

These and other ideas for helping scientists in the new republics during this period of hard transition are in a nascent stage. It is essential that as many as possible come to fruition very soon. □

EC research faces another poor year

■ Maastricht meeting mostly ignores science
■ Pandolfi pushes for increases in Framework

London

FOR Filippo Pandolfi, vice-president of the European Commission in charge of research, 1992 may be another frustrating year. In 1989, he had to accept a 25 per cent cut in the budget for the European Communities' (EC) third five-year Framework research-and-development programme, which was put back a year from its scheduled start date in 1990 and which is still not in full flight. Now Pandolfi faces an uphill struggle if he is to make any real progress towards his stated aim of doubling the EC's spending on research.

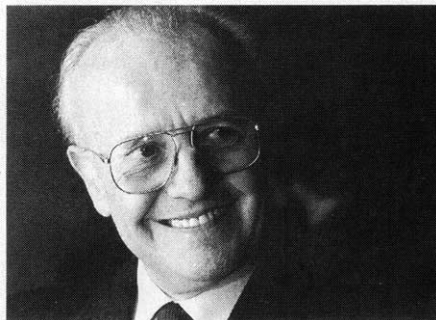
Before December's Maastricht summit, where the EC heads of government thrashed out a treaty under which their 12 separate states will move towards closer economic and political union, Pandolfi had good reason to hope that the principal block to his expansionist plans — Britain's effective veto over increases in EC research spending — would be removed. With UK prime minister John Major fighting to exclude Britain from the other 11 EC states' plans to develop new EC policies on social and employment issues, and anxious to ensure that the Westminster parliament will have the right to decide whether Britain should abandon the pound sterling in favour of a common European currency, Britain was expected to make concessions in lower-profile policy areas — such as research.

But unfortunately for Pandolfi, research and development was so low on the EC leaders' list of priorities that it was raised only in the closing minutes of the two-day Maastricht summit, almost as an afterthought. In a show of hands, Britain alone objected to a proposal that the future EC research budgets should be approved by 'qualified majority' voting (where the larger EC countries have more votes) by the 12 states' research ministers. But with Major having already achieved his two main objectives, none of the other 11 EC leaders had the stomach for another battle with the recalcitrant British. When Pandolfi produces his budget request for the fourth Framework programme (due to start in the mid-1990s) later this year, he will, as in 1989, have to win the approval of all 12 member states — including Britain.

The more money given to the EC for research, argues the Conservative UK government, the more likely it is that some will be used to support product-development work, in conflict with the UK government's policy of reducing its spending on 'near market' research. The vast

majority of the EC research budget already goes to industrial projects, but until now this has been restricted to 'precompetitive' work, where companies work together to develop underlying technologies to strengthen European industry.

The Commission is also allowed to support only those projects that are best carried out through European collaboration, rather than by scientists working in any one member state. Although Britain is not alone in claiming that large increases in EC spending could lead to this rule being broken, the extreme British stance is explained by the rigid accounting practices of the UK Treasury: to a greater extent than in the other EC countries, an increase in British spending on the EC's Framework Programme means a reduction in the amount of money available for research at home.



Pandolfi wants to double research funding.

Pandolfi has two battles on his hands in 1992. By April, he must complete a review of the third Framework programme for the member states' research ministers. Pandolfi is expected to use this opportunity to ask the ministers to restore the 2,000 million ECU (\$2,600 million) removed from the five-year budget at Britain's insistence in 1989. British officials now seem confident that Pandolfi's request will be greeted with little enthusiasm. But some EC-financed researchers, particularly those working to harness energy from nuclear fusion, claim that extra money is needed. The Joint European Torus experiment, based at Culham in Oxfordshire, followed the world's first successful first use of tritium fuel in November 1991 with the announcement that 10 per cent of its staff will lose their jobs. And in 1990, an independent panel headed by Umberto Colombo, now president of the European Science Foundation, recommended that the EC's spending on fusion research until 1994 should be 200 million ECU (\$260 million) higher than the figure allowed under the third Frame-

work budget.

Next autumn, Pandolfi faces the even more formidable task of persuading the research ministers to approve a much-increased budget for the fourth Framework programme. By 1997, he wants to increase the EC's spending on research to around 5,400 million ECU (\$7,000 million) a year — some 6 per cent of the EC's total spending, and more than twice the figure spent in 1991. Pandolfi can make a reasonable case to justify his empire building: the European Parliament backs the plan, and at a summit in Milan back in 1985, so did the EC's member governments. But enthusiasm in the member states for increased EC spending has since waned, as most of Europe wrestles with an economic recession. And with Britain still able to veto the Framework budget if its demands for financial restraint are ignored, Pandolfi will be forced to trim his plans.

The Maastricht summit has also done little to address the interminable delays that have plagued EC research. Although the overall budget for the third Framework was agreed on in a matter of months, the last of its 15 component programmes may not be under way until spring 1992, more than two years later. The problem has been a series of disagreements among the member states' ministers on one hand, and the Commission and the European Parliament on the other, over the programmes' content and management.

Under the Maastricht agreement, approval of the individual research programmes in the fourth Framework should be much quicker — proposals will require only a single cursory reading in the European Parliament, rather than two as at present. But Commission officials predict that any gains there will be wiped out under the new arrangement for agreeing on the total Framework budget. Like individual member states, the European Parliament will also be able to veto the budget proposal. With the Parliament pushing for a large spending increase and keen to flex the limited extra muscle it was given at Maastricht, Commission officials fear a long-running stalemate, with neither the Parliament nor the British government prepared to back down.

Outside the EC, 1991 was a year of belt-tightening for collaborative European projects, and 1992 promises more of the same. The continuing economic recession is one problem, and Germany's role as a leader in European collaboration is compromised by its need to invest in its new eastern states. The biggest question mark hangs over the European Space Agency: Germany is still threatening to pull out of Hermes, the French-led project to build a reusable space plane, if its huge cost cannot be reduced by the addition of new partners (most probably the Japanese) in time for a make-or-break meeting in Madrid later this year. **Peter Aldhous**

Still room for HUGO?

Washington & London

If headlines were what the human genome project was after, 1991 would have been an unqualified success. US genome researcher Craig Venter nearly caused an international incident by filing a surprise patent application for 350 unidentified human gene fragments. Then UK government researchers shocked their US counterparts by trying to charge industry for looking at their genome data.

It was not quite open warfare, but it was twelve months marred by misunderstandings and culture clashes — just the sort of thing, in fact, that genetic pioneer Victor McKusick hoped to avoid when he created the Human Genome Organisation (HUGO) in 1988 to serve as a coordinating body for international genome research. So, as researchers on both sides of the Atlantic crossed signals and missed cues, where was HUGO?

Trying to stay alive, mostly. Three years after its inception, HUGO is "still incipient" in the words of David Galas, the US Department of Energy genome chief. For one thing, it is badly in need of cash. When James Wyngaarden resigned as its director last August, HUGO courted Rockefeller University microbiologist Norman Zinder as a replacement until it realized it could not afford to pay him. Now the position will "remain unfilled indefinitely", says Charles Cantor, vice president of HUGO. "Our resources are really limited."

Nearly all of HUGO's US funding comes from a four-year, \$1-million grant from the Howard Hughes Foundation. In Europe, two British charities, the Wellcome Trust and the Imperial Cancer Research Fund (whose director-general, Sir Walter Bodmer, is HUGO's president) have between them picked up a similar-sized bill.

Many of the problems, Cantor says, stem from early mistakes in HUGO's creation. HUGO is chartered in Switzerland, a move that was supposed to give it tax advantages; instead it turned out to put some strict restrictions on the organization's operations. Swiss law makes it difficult to change articles of federation without long legal processes. When HUGO had to revamp its organization to keep up with the shifting needs of the genome project, the changes turned out to be, as Cantor puts it, "painful and expensive — with a lot of lawyer's fees".

HUGO now has offices in London, Washington, Osaka and Moscow (which opened in July), but only the UK and US offices are incorporated as nonprofit companies, which in the United States entitles HUGO to receive tax-free funds. In Japan, an organization must have a large amount of capital to gain tax-exempt status. So even though HUGO has staffed its Japanese office, it has been caught in a Catch-

22: It does not have enough seed money to win nonprofit status, yet it is difficult to solicit money in Japan until it has that designation.

Even where it has acquired tax-exempt status HUGO has run into problems. A registered nonprofit organization in the United States must by law receive its funding from more than one source, but only a few tiny grants have saved HUGO from being reliant solely on the Hughes Foundation. And because it has no real offices in the United States (except an apartment in Bethesda, Maryland, that Hughes lets it use), it does not have the accounting pro-



Bodmer's HUGO is still finding its place.

cedures and infrastructure necessary to win government grants. Moving under the wing of Johns Hopkins University, something it hopes to do early this year, may finally give HUGO access to the government funding pool.

The Moscow office had barely opened its doors when August's attempted coup threw its future into question. First, Nikolay Laverov, the deputy prime minister who had pledged funds from the Soviet genome project to support HUGO's Moscow venture, was removed from office in the purge that followed the coup's failure. And now that the central Soviet government has collapsed completely, the HUGO office's long-term future is uncertain. Aleksander Bayev, the senior genome project official in Moscow, expects to get money for genome research (and for the HUGO office) from the Russian government, but the 1992 budget has not yet been approved by the Russian parliament.

Despite the difficulties, the Moscow office has made some progress in its two main tasks: to coordinate data collection for the Soviet project, and to improve

communications between Soviet genome researchers and their counterparts in the West (although engineers are still working on an on-line electronic link to the German Cancer Centre in Heidelberg, to give Soviet researchers access to the leading gene databases).

While HUGO has been trying to get up to speed, the genome project has been growing up largely alone — pains and all. In the recent imbroglio over cDNA patenting, for example, the news that Venter had filed a patent for 337 unidentified human gene fragments first came out — together with a flood of controversy — at a HUGO-sponsored workshop during the 11th Human Gene Mapping Workshop in London in August. Some argued that the utter surprise of the UK participants when they learned of the patent application indicated that HUGO has not been doing its job; it was, after all, supposed to be the communications channel for international issues in the genome project. Unilateral decisions about issues as big as gene patenting rarely go over well when they are announced after the fact.

But Cantor points out that the news did nevertheless break at a meeting sponsored by HUGO, not informally or at some later conference. In that sense, although HUGO's efforts at opening communications did not prevent a tumult, they did bring up the issue before things got too far out of hand. Since the meeting, HUGO has set up a committee to study the problem of international data access and make specific recommendations to avoid the problem in the future. Cantor says that he expects that the committee will report in the next several weeks.

Cantor believes that HUGO's darkest days are behind it. Although HUGO was originally set up on what he calls an "elitist academic model" (each of the 400 researchers who have joined HUGO to date have had to be nominated by five existing members, endure a pre-screening process, and be approved by the entire membership), new members will have to get only two nominations and no vote will be necessary. Cantor hopes this will double or triple the membership.

Nevertheless, reinventing HUGO will take more than new rules; for one thing, there is still widespread confusion about just what it is supposed to do. This is partly due to HUGO's changing charter. Three years ago it was hoping to be a great mediator, setting such policies as collaboration and data-trading practices. That lofty aim soon fell by the wayside when it became apparent that, without money, HUGO was not going to be a significant player in genome sweepstakes. Now, Cantor says, "HUGO is primarily a facilitator and a coordinator." HUGO will bring the researchers together, Cantor explains, and then they will set their own policies.

In the coming year, HUGO intends to

take the initiative in organizing the input of data into the main gene-mapping database — the Genome Database (GDB) run from Johns Hopkins. Until now, the 'consensus' mapping data in GDB have been agreed and entered into the database, at the hectic Human Gene Mapping Workshops, held every couple of years. But the volume of data is becoming too large to be handled at a single meeting, and consensus maps for each chromosome will in future be updated at HUGO-sponsored Single Chromosome Workshops, 18 of which are planned for 1992. Once a year, representatives of the teams working on each chromosome will get together to discuss their common problems.

Unfortunately, when HUGO luminaries including Cantor, Bodmer and McKusick outlined the plan to the genome community at the London Human Gene Mapping Workshop in August, many researchers reacted angrily. HUGO offi-

cials in London play down the difficulty as a breakdown in communication: many people assumed that the new meetings would exclude most 'grass roots' genome researchers. In fact, the large biennial meetings will continue (without the time-consuming debates over consensus mapping data, and renamed as Human Genome Mapping Workshops), providing an opportunity for the diverse genome community to get together. This public relations blunder was an inauspicious start to HUGO's new role in gene mapping. The success of the HUGO-sponsored meetings is seen as a litmus test of the organization's ability ever to have an effect on the international coordination of the genome project, and Cantor is depending on a more polished performance in 1992. "The success of HUGO depends on these meetings," he says. "They have to work."

**Christopher Anderson
& Peter Aldhous**

BRITISH RESEARCH FUNDING

Universities face changes

London

BRITISH science leaves 1991 in a more cheerful state than it entered. A year ago, the UK research councils were busy counting pennies, realizing that their funding from government would not keep pace with inflation. The largest of the five, the Science and Engineering Research Council (SERC), which had banked on a much larger budget, was forced to review its whole programme and award only about half the new research grants it had planned. But the research councils' allocations for 1992–93, announced just before Christmas (see table), should allow for a small amount of growth. Even the Agricultural and Food Research Council, which declared a financial emergency in October after the collapse of the UK property market wrecked its plans to sell £7 million worth of empty laboratory buildings, has been given a special £5.7 million 'loan' to ease its cash flow.

With a general election due by the summer, the government's increased generosity is not so surprising. But Britain's researchers should not expect an uneventful 1992, which will see changes in the funding of university research that will dwarf the fallout from last year's research council funding hiccup. The Universities Funding Council (UFC), which distributes nearly £700 million a year of public money to support research (more than twice the amount spent in the universities by the research councils), plans this year to revamp its mechanism for dividing its research budget among the universities. The result will be that money becomes concentrated in the few UK universities that have an outstanding reputation for research; those that can boast little in the

way of research excellence may find themselves suddenly short of funds.

At present, some 40 per cent of the UFC's research funding is distributed simply according to student numbers — universities with more students get more money, irrespective of the quality of the research in their departments. The UFC has been gradually moving towards funding criteria that take a greater stock of the quality of university research, but the government's plans to abolish the distinction between universities and polytechnics in time for the 1993 student entry have forced more rapid progress.

1992-93 budgets for the UK research councils (£million):

Agricultural and Food	107.3	(+8.6%)
Economic and Social	45.1	(+14.3%)
Medical	227.6	(+8.4%)
Natural Environment	129.7	(+2.5%)
Science and Engineering	520.8	(+8.8%)

Figures in brackets give the percentage increase in cash terms from 1991–92 excluding the £47.7 million transferred to the research councils from the UFC.

If the new funding councils that are to fund both the universities and the polytechnics simply adopted the UFC's old funding mechanism, the result would be a sudden redirection of research funds away from the universities and into the polytechnics, which have students galore. (The UFC's counterpart, the Polytechnics and Colleges Funding Council, now spends only several tens of millions of pounds each year on research.) As the government has said that it is not prepared to see this happen, the distribution of research money according to student numbers must end, and to smooth the transition, the UFC

is moving swiftly to bring this about.

UFC officials cannot yet say exactly how the council's 1992–93 research budget will be divided among the universities — the final decision has not yet been taken. But they are confident that the largest factor in deciding each university's allocation will be the UFC's assessment of the quality of research in British university departments, a formidable exercise in peer review that was last attempted in 1989, and will be repeated this year.

The concentration of funding in Britain's leading research universities will be compounded this year by the transfer of some £50 million of UFC money to the research councils, which already direct the majority of their research grants to a select few institutions. (Ten British universities — Birmingham, Cambridge, Edinburgh, Glasgow, Imperial College London, Leeds, Manchester, Oxford, Southampton and University College London — win more than half SERC's university research grants.) This transfer is designed to cover some of the indirect costs of university projects funded by the research councils, and by 1994–95 will be followed by a further £100 million a year that now comes under UFC's budget.

For a fortunate few universities, the changes beginning in 1992 will be welcome. But those with a patchier research record can expect a painful transition. Squeezed from above by the emerging university 'super league', and from below by the polytechnics that wish to claim a slice of the research pie, several university vice-chancellors will face a difficult dilemma over the next few years: whether to divide their shrinking research budget evenly among their departments, or to follow the national trend and penalize weak departments to ensure that research in the strongest is not undermined. The former might minimize faculty protest, but would carry a long-term penalty — any university that takes research money from its best departments to subsidize the rest would soon slip down the UFC's table of research excellence and receive still less money the next time around.

The present turn of events should bring a wry smile to the face of Sir David Phillips, chairman of the Advisory Board for the Research Councils (ABRC). In 1987, the ABRC published its *Strategy for the Science Base*, which proposed that only the best research universities should support a full research programme and said that many UK universities should become teaching-only institutions. After howls of protest from the universities, the government dropped its initial support for the idea. But with the present changes in university funding leading in the much same direction, the ABRC's follow-up to its 1987 document, due to be published in April this year, should make interesting reading.

Peter Aldhous

A final frenzy for landmark cases?

Washington

ANY new year resolution to say no more about misconduct and Robert Gallo, Thereza Imanishi-Kari or David Baltimore will almost certainly be broken. This may, however, be the last year in which that is the case. After years of heated rhetoric and conflicting claims, several landmark cases in the evolution of scientific misconduct policy are nearing their end.

Imanishi-Kari

The 1986 *Cell* paper has been retracted,

co-author David Baltimore has resigned as president of the Rockefeller University, a draft report from National Institutes of Health (NIH) investigators found evidence of misconduct, and yet the case of Thereza Imanishi-Kari and her immunology research could still have a few more scenes to play out.

The US attorney in Baltimore, Maryland, may seek an indictment on criminal charges as early as late January. But now that Suzanne Hadley, the principal investigator in the case, has been forced to

resign from NIH's Office of Scientific Integrity (OSI), the misconduct office is backpedalling as fast as it can from her damning draft report, which was leaked last year. (In fact, OSI no longer even calls it a draft report. Officials now refer to it as the "cross-examination report" and say that the final document in the case will probably bear little resemblance to what has been seen so far.)

But until Imanishi-Kari and her lawyer agree to cooperate, OSI is deadlocked. Imanishi-Kari has refused to comment on

Growing up in public

Washington

If the modern era of scientific misconduct was born two and a half years ago when the National Institutes of Health (NIH) created the Office of Scientific Integrity (OSI), 1991 was its awkward adolescence. The year opened with a nation watching the investigations of AIDS pioneer Robert Gallo and immunologist Thereza Imanishi-Kari (and by extension, her co-author David Baltimore), as well as open warfare over the operation of the OSI. And, unfortunately, it closed just the same way.

In the intervening 12 months, Suzanne Hadley resigned as the deputy director of OSI, and Representative John Dingell (Democrat, Michigan) strongly criticized NIH for its bungled handling of the whole issue. Other than that, not much changed. Investigation of scientific misconduct was a mess last year, and it is a mess today.

However, 1992 may be the year in which misconduct grows up. For one thing, the investigations of Imanishi-Kari and Gallo — OSI's flagship cases — seem to be winding down, though slowly (see story this page). And although those cases have been long, ugly affairs, they have opened up the misconduct system as never before.

Through congressional hearings, a phenomenal amount of news coverage, and the attention of virtually every element of the scientific community, the pitfalls of misconduct investigating are now a matter of public record. Leaks are one problem. So are inconsistent procedures (for instance, prominent researchers got special review committees, although others did not). In both the Imanishi-Kari and the Gallo case, NIH investigators were often reduced to a role of following up allegations in the press, which made nearly everyone but Dingell uncomfortable. And an important debate over 'due process' in OSI investigations has, intentionally or not, essentially halted several cases.

Even OSI admits that some of its most prominent investigations were badly handled. But it has also learned some tricks on the job: to avoid leaks, sensitive drafts reports now go only to principal parties, and OSI is increasingly employing forensic and statistical analysis to add some quantitative rigour to what has often been a disquietingly subjective process. Investigations now focus on whether misconduct occurred, and no longer stumble on the question of a researcher's intent. As OSI discovered, claims of "unintentional" misconduct have flummoxed many university investigations, even when they turned out to be a red herrings that obscured clear abrogation of scientific responsibilities.

Other changes at OSI are coming from outside. After losing a lawsuit that challenged the way it developed its procedures, OSI published a set of proposed new rules last year. Public comments were generally scathing, mostly focused on the

proposed definition of misconduct, which included, together with the usual "fabrication, falsification and plagiarism", the category of "other practices that seriously deviate from those that are commonly accepted from the scientific community". An NIH advisory committee has recommended that the catch-all phrase be changed to "other fraudulent activities in proposing, conducting, reporting or reviewing research", a definition that OSI says it can live with.

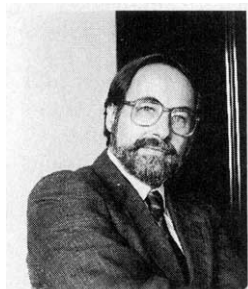
The committee also proposed — and NIH agreed — that OSI's staff be increased from 19 to 28, including, for the first time, three lawyers (OSI investigators have traditionally been scientists). And the committee recommended open hearings, in which accused and accuser can face each other. OSI director Jules Hallum opposes that move, arguing that face-to-face confrontations "would destroy the willingness of whistle-blowers to come forward."

Even as it reconsiders its role, however, OSI languishes in a sort of bureaucratic limbo. Both Congress and some Administration officials are contemplating taking OSI away from NIH and placing it instead under the wing of the Department of Health and Human Services, NIH's parent agency. When Dingell held a hearing last summer accusing Bernadine Healy, the NIH director, of a conflict of interest in an OSI investigation of a case at the Cleveland Clinic, Healy's former institution, it only reinforced the concern that OSI — located on a campus full of scientists — is vulnerable to pressures from the scientific community. Dingell thinks OSI might be more independent if it operated like any other government investigative office — firmly entrenched in the bureaucracy. If the administration does not propose the move itself, congressional legislation to that effect may appear this year.

But the worst may be behind the misconduct controversy, if not OSI itself. Perhaps the most encouraging sign is the improving quality of university investigations. Whereas academic panels in the past often erred on the side of finding no misconduct, Hallum says that recent university investigations, such as two last year at the California Institute of Technology, have been more thorough and fair. "If they keep it up, they may put us out of business," he says. Nevertheless, until conspicuous mishandlings such as Imanishi-Kari's inquiry at the Massachusetts Institute of Technology and that of whistle-blower Erdem Cantekin at the University of Pittsburgh (see story, next page) become a thing of the past, OSI wants to keep tight reins on the universities. The proposed new procedures would allow OSI to intercede earlier in an academic investigation if things seem to be going awry, and Hallum is hoping to have the rules clarified to give OSI explicit authorization to investigate the universities themselves, to explore the possibilities of cover-ups.

Christopher Anderson

the draft report (beyond an attack on the overall process) until she is allowed to see the original laboratory notebooks on which OSI and other federal investigators carried out the forensic analysis that led them



David Baltimore

to conclude that data had been fabricated. Many of those notebooks, however, are in the hands of the US attorney, who has declined OSI's repeated request to turn them over to Imanishi-Kari. If a grand jury

reaches an indictment, a legal process known as 'discovery' will allow Imanishi-Kari and her lawyer to examine the data. Until then, OSI considers its hands tied.

The congressional Investigations and Oversight Committee of Representative John Dingell (Democrat, Michigan), which has held four hearings on the case since 1988, may hold its last in the spring — a 'lessons learned' hearing to review the way in which NIH and the relevant universities handled the allegations.

Gallo

AIDS pioneer Robert Gallo had a stormy year in 1991, and 1992 may be almost as bad, but at least one cloud should soon clear. Sources close to the investigation say that OSI is nearing completion of a report that should essentially clear Gallo of scientific misconduct in identifying and claiming the AIDS virus on the basis of a sample that later turned out to be an isolate from French scientists. In its place a new dispute is likely to arise over the 1987 patent on the AIDS test, which, by agreement, is shared by the United States and France. Central to the controversy is the question of whether Gallo — and by extension, US government officials — knew at the time of the patent application that the virus had actually been first isolated by the French.

French government officials and their lawyers are lobbying their US counterparts to reopen the patent agree-

ments, on the grounds that they were clearly based on assumptions that have turned out to be untrue. The 1987 agreement allowed for that prospect, but not for the possibility that the statements were known to false at the time they were made. Dingell's staff, who have closely followed the case, are pursuing allegations that Gallo knew



Robert Gallo

about a 1984 study conducted by Donald Francis of the US Centers for Disease Control that showed that the US and French blood tests performed at about the same level, something that would have indicated that the French had indeed isolated the real AIDS virus.

There is little doubt that a study was performed. What Dingell must prove is that Gallo was aware of the study's conclusion at the time, and that the study did show convincingly that the blood tests performed similarly. Joseph Onek, Gallo's lawyer, says that he has not seen such a study and knows of "absolutely no evidence to support the charges" that Gallo should have known that the tests were equivalent. Indeed, he says, even Luc Montagnier, Gallo's French competitor, was emphasizing differences between the viruses at the time. If experience is any guide, resolving this issue could take all year.

Cantekin

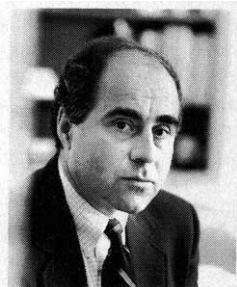
A less well-known case, but one just as relevant to the current controversy over scientific misconduct, is that of Erdem Cantekin, a bioengineer at the University of Pittsburgh.

Last month, the *Journal of the*

American Medical Association published a long-delayed paper by Cantekin accompanied by an unusual five-page commentary by the West Coast editor, **D r u m m o n d**

Rennie, explaining the paper's convoluted history. Five years in the making, the story has emerged in bits and pieces. Essentially, Cantekin and Charles Bluestone, another University of Pittsburgh researcher, had collaborated on a clinical trial of an antibiotic known as amoxicillin on ear infections in children. Bluestone concluded that the drug was effective; Cantekin disagreed. In 1986, the two submitted contradictory articles to the *New England Journal of Medicine*, which asked the university to identify the 'authorized' principal investigator. Pittsburgh identified Bluestone and then began disciplinary action against Cantekin for attempting to take unwarranted credit for the study.

Although an OSI investigation cleared Cantekin of wrongdoing, Pittsburgh has continued its attempts to strip him of tenure. Now that his dissenting paper has finally been published, Cantekin is preparing to take the offensive. He has filed a lawsuit against Bluestone and the hospital where they did the study to attempt to gain possession of the original data, and has



Erdem Cantekin

served six current and former Pittsburgh officials with a 40-page complaint and intent to sue, charging that they participated in a conspiracy to withhold data from the public and damage his career. He also plans to petition the US Food and Drug Agency to withdraw amoxicillin from use in children's ear infections, something that would effectively halt a market estimated at 30 million prescription a year.

Christopher Anderson

SPACE SCIENCE

Smaller, cheaper, leaner

Washington

If funding for the US National Aeronautics and Space Administration (NASA) keeps up with inflation over the next few years, officials will count themselves fortunate. Throughout the agency, planners are rediscovering austerity, and shedding grand space plans as fast as they developed them in the boom years of the late 1980s.

After Congress in 1991 rejected NASA's proposal to start planning for a mission to Mars, the agency is planning to return with a much more modest strategy involving cheap robotic missions. In its 1993 budget request (which will be released next month) NASA is expected to ask for some \$50 million in start-up funds to begin planning several unmanned missions to the Moon, with the aim of developing techniques that could be used someday on Mars.

The proposed space station, which spent much of 1991 precariously balanced on the edge of cancellation, now seems politically secure, if somewhat battered. Still overweight, underpowered, and needing constant maintenance in its latest planned incarnation, the project has become an mechanical, rather than a political, challenge. "It's up to the engineers to make it work now," says George Washington University space analyst John Logsdon.

But if the space station will not be 1992's whipping boy, the Advanced Solid Rocket Motor may take its place. The \$465-million effort to develop a next generation of engines for the space shuttle owes more to the desire of House Appropriations Committee chairman Jamie Whitten (Democrat, Mississippi) to create jobs in his home state than any pressing NASA needs. Now that the shuttle programme itself has been discontinued, the Administration is planning to put its foot down on what it sees as a particularly egregious example of 'pork barrel' funding. President Bush is expected to seek the cancellation of the project this year, something that seems sure to prompt a nasty response from Whitten and could tangle the rest of the NASA budget in political infighting.

Christopher Anderson

Japan leaning towards SSC

Tokyo

As Japan heads into 1992, it faces two important, related issues: the increasingly urgent calls for reform and renovation of its university research system, and the question of how to respond to US requests for help in building the Superconducting Super Collider (SSC).

On the second issue, a preliminary decision seems to have been made. It now seems almost certain that Prime Minister Kiichi Miyazawa will promise financial support for the SSC when US President George Bush visits Japan next week (7–10 January). Miyazawa's message of support will come despite widespread opposition in the Japanese government bureaucracy and academic institutions to helping support the SSC. The SSC is now estimated to cost \$8,500 million, and the US Administration is seeking about \$1,000 million from Japan.

Officials of Japan's science-related ministries and agencies have made it plain to senior US officials who have visited Japan in the past few months that Japanese government organizations cannot afford to contribute to the SSC out of their limited budgets. And government officials have repeatedly explained to their US counterparts that the priority for Japan at present is domestic reform and renovation of the university research system.

Nevertheless, Miyazawa seems determined to press ahead with support for the

SSC. In mid-December, the Miyazawa administration made an abortive attempt to introduce taxes in 1992 to fund contributions to international activities, including Japanese participation in the SSC. But the proposal to raise ¥500,000 million (\$3,800 million) in extra taxes ran into a wall of opposition from industry and even from the ranks of Miyazawa's own ruling Liberal Democratic Party (LDP), partly because it included an attempt to continue temporary taxes on petroleum products and cars. The temporary taxes were introduced to help fund Japan's contribution to the Gulf War on the understanding that they would last only one year.

Undaunted, Miyazawa now says funds for the SSC will somehow be included in the 1993 budget. But just how this will be done without cutting into other budgets, no one knows.

While Miyazawa was trying to raise hundreds of millions of dollars for the SSC, officials of Japan's Ministry of Education, Science and Culture (Monbusho) were burning the midnight oil last week in last-minute negotiations with the Ministry of Finance trying to obtain about \$100 million in fiscal year 1992 to start renovating Japan's universities. Pressure from industry and universities to reform and rebuild Japan's universities — in particular, Tokyo University — has reached a critical point, and the ministry of education has at last started to request extra

funds for the university research system.

A powerful new science pressure group has recently been formed within the LDP headed by Kishiro Nakamura, a former head of the Science and Technology Agency, to push forward renovation of the national laboratories and universities and the internationalization of Japan's science. On 26 December, Miyazawa told Nakamura that he expects that Bush will request the formation of a US-Japan working group on the SSC and Miyazawa would like to cooperate with the United States.

However, the position of Nakamura's committee on the SSC is equivocal. On 19 December the committee issued a statement saying that the collider will be the most advanced facility in the world, that it will contribute to fundamental understanding of the Universe and materials, and that how Japan deals with the SSC issue will have a fundamental bearing on the future of Japan's global partnership with the United States. But the committee then went on to say that to achieve the right environment for cooperation on the SSC and other international projects, the science-related ministries and agencies must first give urgent priority to improving the research base in the universities and national research laboratories.

No matter what the committee's position is, most science-related officials feel that Miyazawa will make some promise of support for the SSC, but it seems unlikely that he will be able to put a dollar figure on the Japanese contribution or say where it will come from. **David Swinbanks**

Large Hadron Collider moves ahead

London

As the United States tries to find help paying for the SSC (see above), plans to build a similar machine at CERN, the European particle physics laboratory in Geneva, are gathering pace. The Large Hadron Collider (LHC) would be much less expensive than the SSC — around SFr2,000 million (about \$1,400 million) — because it will be built in the tunnel housing CERN's existing Large Electron-Positron collider. Director-general Carlo Rubbia says he is confident that the LHC can be built without increasing CERN's total budget, but the two detectors planned for the machine must be paid for separately, and are expected to cost at least another SFr600 million (about \$425 million).

In Germany and Britain, the cost of the LHC has been causing some anxiety. Before a meeting at CERN in late December, where Rubbia expounded the LHC's merits to a gathering of senior European science policy makers, there was concern that the German and British doubts would delay progress. But Rubbia is now

in an ebullient mood, having won the backing from the CERN member states to continue with research on the LHC's magnets, and to choose the detectors for the machine.

At the insistence of Britain and Germany, Rubbia must produce acceptable blueprints for the detectors before winning final approval for the LHC, which will defer that decision until late 1993. But, given the problems experienced at SSC with the selection of detectors (which was left until after the main design had been approved), Rubbia is happy with the timetable he has been given. He says it is important to develop a collider and detectors that together will produce good science, rather than first designing a machine that will "go into the *Guinness Book of Records*" and then thinking about the programme of experiments. "Our friends across the Atlantic neglected this," Rubbia says.

Work on choosing the detectors for the LHC will begin in March, at a meeting in Evian-les-Bains in France. At present there are four designs competing for two detec-

tor slots, and CERN's management hopes that the issue can be decided without the controversy that has surrounded the SSC's detectors.

Rubbia has also been asked to trawl for financial contributions towards the LHC from countries that are not CERN members. He believes he has a good case, as physicists from non-member states make up a quarter of those working at the laboratory. "We're totally overbooked, and many people from the member states are complaining," says Rubbia, suggesting that non-member countries may find it difficult to obtain time for their physicists at CERN in future, if they do not contribute towards the LHC. Rubbia does not intend to ask the United States for donations — he is happy with the reciprocal exchange of scientists between CERN and the United States, and says there is no need for money to change hands — but most other countries whose physicists use CERN can expect to hear from Rubbia over the coming year.

Peter Aldhous

Looking expectantly to Rio

London

ENVIRONMENTAL issues will return to the centre of the international stage in June, when Rio de Janeiro hosts the United Nations Conference on Environment and Development. The summit should see the signing of legally binding international treaties to address two of the world's most pressing environmental problems: the threat of global warming and the erosion of the world's biological diversity.

Since February 1991, negotiators have been struggling to draft a climate change convention that can satisfy all the world's governments. The main stumbling block has been the deep rift between the European Communities (EC) and the United States over the question of setting a target to limit carbon dioxide emissions from the developed world.

A two-week meeting in Geneva that ended on 20 December did little to resolve the impasse and, with only one more negotiating session scheduled before June, the convention seems certain to contain no more than a legal framework under which subsequent action to counter global warming will be agreed — rather like the 1985 Vienna Convention, which preceded the Montreal Protocol on the protection of the ozone layer. A question mark also hangs over the involvement of the former Soviet Union. This is not just an academic question: in 1991, the Soviet Union was still the world's largest producer of oil, even after a 25 per cent decline in production since the late 1980s. The key factor will be the emerging policy of Boris Yeltsin's Russian government, which controls more than 80 per cent of the former Soviet Union's oil reserves — although it is safe to assume that climate change will not be among Yeltsin's priorities.

Despite disappointment among European governments about the expected failure to draft a climate convention with any real teeth, even the bland document likely to be signed in Rio will commit the world's governments to continued negotiations. And there are some initial signs of a softening attitude from the US Administration. (The departure of White House chief of staff John Sununu, seen as the principal architect of President George Bush's climate change policy, may be an important factor.) At the December meeting in Geneva, the US negotiator, Robert Reinstein, announced that the United States will in 1993 produce a detailed plan of US action to combat global warming (at the federal and state government level and in the private sector). Significantly, Reinstein said that the plan could include some new measures.

Like the climate treaty, the convention on biodiversity is expected to be little more than a bare legal framework, under which future conservation measures will be negotiated. But although environmentalists are disappointed with the progress on climate, those attending the parallel talks on biodiversity are relieved that any progress has been made at all.

Four months ago, says Simon Lyster from the World Wide Fund for Nature, the gulf between the developed and developing countries was so large that the prospect of drafting any treaty in time for the Earth Summit seemed remote. There are still big points of disagreement — for example, over the patenting of biotechnology developed in the industrialized countries but using genes from the developing world — but both camps have begun to compromise.

Peter Aldhous

Pinning the blame on carbon dioxide

London

THE growing consensus among climatologists is that cutting emissions of carbon dioxide, rather than further controls on other greenhouse gases, is the surest way to slow any global warming caused by the greenhouse effect.

In October, atmospheric scientists asked by the United Nations Environment Programme to assess the damage to the stratospheric ozone layer announced that chlorofluorocarbons (CFCs) may cause less warming than was thought: CFCs help to trap heat in the Earth's atmosphere, but the destruction of ozone, at least regionally, can have an opposite and roughly equal effect. Although phasing out CFCs is still necessary if the ozone layer is to repair itself, their demise will not curb global warming

as much as had been thought.

The atmospheric chemistry of methane and nitrogen oxides is now also under close scrutiny. Both react with hydroxyl radicals to produce gases that can affect climate — carbon dioxide and water vapour in the case of methane, and tropospheric ozone from nitrogen oxides. Again, climatologists now have less confidence in their calculations of these 'indirect' influences on climate.

The narrowing focus on carbon dioxide as the principal culprit in climate change is likely to weaken the US negotiating position in the Rio de Janeiro Earth Summit. The US Administration has insisted that its climate change policy should be judged by looking at US controls on all greenhouse gases, rather than just carbon dioxide. P.A.

New research centres

Sydney

THE Australian government has announced the formation of 20 new research centres, a move that will eventually inject up to A\$100 million (\$77 million) into Australian science each year.

The 20 centres represent the second and largest step in the government's Co-operative Research Centre (CRC) programme — a principal part of its effort to boost Australia's advanced industrial base through pure and applied research. As with the first 15 centres of the programme announced last March (50 are planned), the government will grant about \$2 million a year to each centre, provided that the centre partners put in at least the same amount in cash or kind.

After the new centres become fully operational by 1993–94, the government will provide \$250 million of the centre's total seven-year budget of \$764 million. Business will contribute about 16 per cent, with the rest coming from universities, federal and state departments and agencies and medical research institutions.

Almost the only rule in the programme is that each centre must have a university as a partner. However, the research centres, which will be chosen by a special committee of scientists reporting to the Minister for Science, Ross Free, do emphasize certain themes, including materials research, communications and information technology, medical research and molecular engineering, mineral exploration and ore refining, and agriculture. One centre, the Australian Maritime Engineering CRC, will help Australia's fledgling shipbuilding industry. Mark Lawson

Woman to head CSIRO

Sydney

THE participation of women in Australian science has been given a boost by the appointment of Professor Adrienne Clarke as chairman of the nation's premier research organization.

Clarke, a molecular biologist, was recently appointed chairman of the Commonwealth Science and Industrial Research Organization (CSIRO), replacing a former senior state politician, Neville Wran. As the CSIRO is by far the largest and most influential Australian research organization, the appointment makes Clarke one of the nation's most senior scientists.

Clarke, who has been a CSIRO board member since 1986, is head of the University of Melbourne's School of Botany and director of the university's Plant Cell Biology Research Centre.

Her first act as chairman was to announce an extra A\$3 million for research into an algal bloom infesting a southeastern river system. Mark Lawson

Olympic row over sex testing

SIR — Your News item about sex testing at the Olympics (*Nature* 353, 784; 1991) raises an important issue that has caused serious concern among medical geneticists, endocrinologists and others for more than 20 years. The problem stems from the decision taken by the International Olympic Committee (IOC) in 1967 to use sex chromatin testing as a criterion of eligibility for women competitors. The purpose was to exclude males and people with intersex conditions, who might have an unfair advantage in muscle mass, from competing in women's events. Its use since 1968 has helped to reduce innuendo and scandal in the mass media about individual competitors who were thought to have a 'masculine' body habitus.

But there is no evidence that its use has led to the detection of males masquerading in female events. On the contrary, and as predicted by medical geneticists, a substantial number of women athletes have been unfairly debarred from competing. Although the IOC has not responded to requests for information it holds on the number of athletes found on testing to be ineligible, it is believed that at least 1 in 500 female athletes have had to withdraw from competition on this account (M. A. Ferguson-Smith, & E. A. Ferris, *Br. J. Sports Med.* 25, 17-21; 1991). As your article correctly states, the majority of women affected are "XY females" with the androgen insensitivity syndrome who clearly have no athletic advantage over normal XX women.

It is not hard to imagine the distress and humiliation felt by an athlete confronted with failure to pass a sex test. A graphic account of the trials of one courageous athlete who found herself in this position can be found in the March 1991 issue of *Women's Sports & Fitness*. In brief, the athlete passed the test in 1983, forgot her eligibility certificate in 1985 and failed the test when it was repeated. There followed a harrowing period during which she was publicly debarred from competition, lost her scholarship and had her achievements struck from the record books. Finally her persistence was rewarded and she was reinstated in 1988. There are many cases where an athlete confronted with a similar experience has preferred to withdraw rather than submit to further humiliation.

The time is long since past for these inappropriate sex tests to be abandoned. However, the IOC seem unconcerned and proposes to introduce a polymerase chain reaction test for Y-specific DNA for all female competitors at the forthcoming Olympic Games in Albertville

and Barcelona. The Council of the International Amateur Athletic Federation has a more sympathetic approach. It wishes to see genetic sex tests dropped and has recently made a quite different proposal. It believes that there are good medical reasons for all athletes, men and women, to have a health check before competing to exclude physical disorders that might put the athlete at risk (for example, Marfan's syndrome among basketball players). Although not primarily designed for gender verification, such an examination would obviously not allow a male to masquerade as a female. The health check would be organized in the athlete's own country by sports physicians accredited by the National Olympic Committee. Random checks for gender verification would not be required at international events, as the procedures used for the collection of urine samples during control of anabolic steroids and other drugs would readily identify males. These arrangements should provide sufficient assurance to the IOC and those women athletes who require it that all competitors are eligible. To persist with genetic sex tests would simply continue the present discrimination against women athletes.

M. A. FERGUSON-SMITH
(Department of Pathology, University of Cambridge, Cambridge, UK), ALISON CARLSON (Amberst Drive, Hasting-on-Hudson, New York, USA), ALBERT DE LA CHAPPELLE (Department of Medical Genetics, University of Helsinki, Finland), ANKE EHRHARDT (Department of Psychiatry, Columbia University, New York, USA), ELIZABETH FERRIS (Brown Candover, near Aylesford, Hants, UK), ARNE LJUNGQVIST (Department of Pathology, Karolinska Institute Stockholm, Sweden), MYRON GENEL (Yale University School of Medicine, New Haven, Connecticut, USA), JOE LEIGH SIMPSON (University of Tennessee College of Medicine, Memphis, Tennessee USA)

Future fuels

SIR — Hall *et al.*¹, in their Commentary article on the greenhouse advantages of using solar energy via biofuels, seem to have overlooked some realities. First, barring a vast catastrophe, the human population will reach or exceed 10,000 million by the next century. Second, there is a global food crisis. It is clear that nutrition is not simply calories, and the Sadik report² shows the scale of the problem, with endemic anaemia in Egypt, India, Indonesia and Pakistan, and so on. Food per capita is declining in about half the world, even in biofuelled Brazil. Third, global soil erosion is

reaching crisis proportions^{3,4} and reliable water resources are a problem in many regions.

Yes, there are lands such as United States, Canada and parts of Europe that could move to massive use of biofuels. They all have the necessary conditions of a land-food surplus. But such countries are also the emergency bread-baskets of the world, and who would hold the surplus if they diverted possible resources to biofuels?

Climate is not steady-state on many timescales. In 1988, the United States had a corn deficit. What would happen to food and fuel if we had a year without summer, perhaps driven by a really large volcanic event⁵? We have not had such an event this century.

At present, we spend very large sums of money and expend much energy on the development of nuclear fusion energy. But we have access to a very nice fusion reactor in the Sun, which delivers about 2 cal cm⁻² min⁻¹. Surely this must be the sustainable energy source for the future, with the great deserts used as energy farms, and every rooftop and wind source used on local and regional scales? Solar energy systems, combined with rational population reduction, could lead to a planet of diversity and opportunity.

W. S. FYFE

Department of Geology,
University of Western Ontario,
London, Canada N6A 5B7

1. Hall, D. O., Mynick, H. E. & Williams, R. H. *Nature* 353, 11-12 (1991).
2. Sadik, N. *State of World Population 1989* (UNFPA, 1989).
3. Brown, L. R. *State of the World 1991* (Norton, 1991).
4. Fyfe, W. S. *Episodes* 12, 249-254 (1989).
5. Grove, J. M. *The Little Ice Age* (Methuen, 1988).

Infant food

SIR — As Nigerians, we suggest that the answer to the issue of advertising campaigns for formula feeding in developing countries lies somewhere between the view expressed by Jenny Daniel (*Nature* 353, 496; 1991) and that of the author of your leading article (*Nature* 352, 266; 1991). Manufacturers of formula feeds have in recent years been responsive to criticisms by the World Health Organisation (WHO) and other international bodies, but clearly there is room for improvement. It would seem best to us, however, that international agencies and governments as well as manufacturers of formula feeds should spend more effort on emphasizing the virtues of breastfeeding, thus enabling those of us in developing countries to make both a free and informed choice based on the facts.

A. O. ADEBAJO
A. F. ADEBAJO

University of Ibadan,
Nigeria

Eighteen-ninety two and all that

J. L. Heilbron and W. F. Bynum

1992's anniversarial celebration encompasses the Old World and the New, by route of Tasmania. The Celsius scale and the revolutionary calendar join other births (Babbage) and deaths (Schrapnel) in this year's cornucopia.

THIS year five centuries ago Columbus took his trip to the Bahamas. He did not, of course, discover America: he met many people there who knew about it before he did. Nor, probably, was he the first European to see the New World. The latest news from the archives adds two travelling salesmen from Bristol to the list of those who may have arrived in Newfoundland or New England before 12 October 1492. Still, we take the observance of the quincentennial of Columbus' landfall as our anniversary of the year.

Columbus may not have discovered America; but he certainly discovered a reliable way there and back, and, in contrast to earlier visitors, publicized his works and deeds. Also, presumably again in contrast to his predecessors, he knew precisely what he sought and believed that he had found it. In that he was mistaken. Columbus thought that he had reached the Indies because he hit land where he reckoned they were located, about 2,000 nautical miles, or 30 days' sail, west of the Canary Islands. All the competent advisors to all the kings he had approached knew the distance to be much greater: the famous dispute about the possibility of the voyage did not oppose Columbus to flat-earthers, but a cock-eyed calculation, off by a factor of five, to a balanced consideration of the evidence. Columbus deserves to be celebrated for his stubbornness, conviction and bravery, and for showing that perseverance in error is no bar to success. The accomplishment for which the anniversary of the year will be celebrated to surfeit may best be specified in the impenetrable logic of Mark Twain's Pudd'nhead Wilson: "It was wonderful that Columbus discovered America; it would have been more wonderful still if he had missed it."

There is another claim to the anniversary of the year. In 1642, Galileo died and Newton was born. The obvious inference, that the soul of the one passed to the other, has not withstood our scrutiny. Galileo died on the Gregorian calendar and Newton was born on the Julian. Had Galileo been alive when Newton entered the world on 25 December 1642, English style, he would have made the date 4 January 1643; had Newton been alive and counting when Galileo died on 8 January 1642, Italian style, he would have made it 29 Decem-

ber 1641. Turn it as you will, they did not come to be and pass away in the same year; hence, no transmigration of souls, and no distinguished anniversary. Further insights into calendrics are supplied below, at 1792, 1742 and 1642.

In keeping with our earlier format, we begin with centennials (1.0¢), advance to sesquicentennials (1.5¢), and continue at 50-year intervals back to 1492. We end with our century, with the semi-centennial of an event that returns us to the anniversary of the year. We also call attention to an innovation in our

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Anniversarian of the year.

art, the average or weighted-average anniversary, examples of which are given under 1892, 1842 and 1592.

1892 (1.0 centenary)

The discoveries in the sciences of 1892 as, doubtless, of most years, offer a choice between compulsive detail and soaring speculation. We have, on the one side, the detection of Amalthea, the fifth moon of Jupiter and the first to be found since Galileo spied the Medici stars in 1610. The discoverer, Edward Emerson Bernard, had keen eyes: the shy Amalthea shines at the thirteenth magnitude. Further to detail, Albert Abraham Michelson (born 1852, 1.4¢) began in 1892 his measurement of the standard international meter bar in terms of the wavelength of the red cadmium line; a work he concluded one year and 1,553,163.5 wavelengths later. He thereby made otiose the result obtained by the original measurers of the meter, who had begun their lengthy labours exactly a century earlier (see

under 1792). Michelson could have made good use of the adding machine invented, also in 1892, by his compatriot William S. Burroughs. Further to compulsion, Francis Galton (born 1822, 1.7¢), whose grand passion was to quantify and systematize, issued 100 years ago his version of that essential guide to modern living, the classification of fingerprints. One of Galton's admirers, August Weismann, sent him a copy of *Das Keimplasma*, published that year. Although Weismann's researches did much to integrate Galton's notions of hard heredity with the latest microscopic investigations of cells, it is not at all clear that Galton understood the nuances of Weismann's present.

On the soaring speculative side we have Hendrik Antoon Lorentz, who published his first account of the electrodynamics of moving bodies in 1892. The account included the suggestion that the failure of the aforementioned Michelson and his collaborator Edward Williams Morley to detect the motion of the Earth through the ether arose from a contraction of their apparatus in the direction of the supposed motion. The failure and the suggestion may have played a part in the creation of the theory of relativity. Another high soarer was Alfred Werner, who invented coordination chemistry between 1891 and 1893; we assign it the weighted-average date of 1892. The definitive theory, which broke with previous ideas about chemical bonding, ascribed to metal ions a primary valence, satisfied only by ions, and a secondary valence, saturated by a fixed number (the 'coordination number') of neutral molecules.

Further to chemistry, we may notice two substances that, during the past century, have come to be identified as public nuisances. In 1892 A. Kühlewein introduced asbestos cement, a very effective glue and caulk now banned in California; and A. Pinner worked out the structure of nicotine, which is doing only a little better than asbestos. In keeping with their reputation as one of the less desirable of New World imports, tobacco plants themselves suffer illnesses. One of these, tobacco mosaic disease, was investigated from 1890 by Dmitri Iosifovich Ivanovsky. Beginning his work with the common assumption that the causative agent was a bacterium, he tried the usual filtering methods to

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Determining atmospheric pressure.

isolate it, but discovered instead that the filtered sap still contained the infective agent. Hence *filtrable virus*, 'virus' then meaning any agent that causes disease. Ivanovsky published his results in 1892 and so 'discovered' viruses.

Chemistry lost some important practitioners in 1892: Archibald Scott Couper, who introduced structural principles of organic chemistry, independently of Kekulé; August Wilhelm von Hofmann, first director of the Royal College of Chemistry best known for his work on coal tar; and Hermann Kopp, who enjoyed the tedious quantifications, and laborious measurements, needed to obtain specific heats and gravities, all died. So did John Couch Adams, co-discoverer of the planet Neptune; George Biddell Airy, long-time Astronomer Royal of England; Richard Owen, whose soaring reputation was eclipsed by Darwin; and Werner von Siemens, an artillery officer turned electrical inventor and manufacturer.

On the credit side, J. B. S. Haldane, who studied genes, enzymes and Indian philosophy, Louis de Broglie, who gave the impetus to wave mechanics, and Arthur Holly Compton, whose demonstration, in 1922 (0.7¢), of the particulate character of X-rays, which inspired de Broglie, were born; as was the University of Chicago, one of the earliest research universities in the United States, where Compton worked from 1923 to 1945.

1842 (1.5 centenary)

Our sesquicentennial should resound. In 1842, Johann Christian Doppler announced his explanation of the rise in pitch of a sound as its source moves towards us, and its drop in pitch as its source recedes. The same year saw the

Mary Evans

birth of the greatest theorist of sound since Pythagoras, John William Strutt, third baron Rayleigh. Rayleigh also deserves a centennial notice. In 1892 (1.0¢), he pointed out that nitrogen obtained from the atmosphere weighed more than nitrogen obtained from chemical compounds, an observation that, three years later, resulted in the discovery of argon by himself and William Ramsey (born 1852, 1.4¢). On a quieter note, Crawford Long in rural Georgia first used ether as a surgical anaesthetic. The patient he operated on was asleep and Long did not shout his achievement from the housetops, so the noisy noticing of anaesthesia's 1.5¢ birthday must wait another four years.

For chemists, 1842 marks the publication of Justus von Liebig's famous treatise on animal physiology, *Die Tierchemie oder die Organische Chemie in ihrer Anwendung auf Physiologie und Pathologie*. It is also the average date of Carl Gustav Mosander's separation of didymia from lanthanum and terbium and erbium from gadolinite, and of Eugène Melchior Péligat's isolation of uranium metal, which occurred precisely 100 years before Enrico Fermi used it to charge up the world's first radioactive pile (see under 1942).

We are pleased to report the death of Henry Schrapnel, the inventor of the murderous shell named after him, and the births of James Dewar, of the flask; of Camille Flammarion, astronomer and popularizer, whose *Les Merveilles Célestes, la Planète Mars et ses Conditions d'Habileté*, published in 1892 (1.0¢), titillated hunters for extra-terrestrial life at the turn of the century; and of William James, psychologist, author of *The Varieties of Religious Experience* (1902, 0.9¢) and *The Principles of Psychology*, the centennial of which was noted two years ago.

1792 (2.0 centenary)

For the further perplexing of the calendar, the revolutionary government of France declared a new era beginning with the day of the autumnal equinox of 1792. The first revolutionary year thus included about 100 days from what kings and tyrants reckoned as 1792, and the balance from the oppressors' 1793; and, as a greater aid to democracy, the Republic divided the day into 10 hours and the hour into 100 revolutionary minutes. The era lasted until New Year's day, 1806, when Napoleon put an end to it.

The French adopted a new telegraph as well as a new calendar in 1792. The invention of Claude Chappe, it consisted of a post with a transverse bar at the ends of which two smaller arms turned on pivots. A code transcribed various orientations of the arms into a message; a series of telegraphs within view of one

another provided for its propagation. Chappe's telegraph might have transmitted news of the publication of Arthur Young's *Travels in France*, in which the greatest of England's enlightened writers on agriculture criticized the French for their poor methods of working their rich soil; and of the setting out of Pierre-François-André Méchain and Jean-Baptiste Joseph Delambre (died 1822, 1.7¢) on their mission to find the length of the meridian through Paris, from the Gulf of Lyon to the English Channel. Their purpose was to establish a permanent and natural basis for a new system of weights and measures for the civilized peoples of Europe.

Their labours on their meridian made no difference to George Vancouver, who had left England in 1791 on the voyage that would take him, via the Cape of Good Hope and the Hawaiian Islands, to North America. There, in 1792, he carefully surveyed the coasts from California up to Canada, discovering the Gulf of Georgia and circumnavigating Vancouver Island.

We must signal the birth of Charles Babbage, like Burroughs (1892) and Pascal (1642) the inventor of the calculating machine; and also a strong force in the strengthening of British science to compete with the powerful school of French mathematical physicists represented by Laplace, Fourier and Ampère. Another numerate British *savant*, J. F. W. Herschel, first saw the light in 1792; though some of his later sightings of luminous heavenly bodies (double stars, the Eta Carinae nebula) received more acclaim. While his good friend Babbage ranted about "the decline of British science" (history is cyclical) in the 1820s, Herschel narrowly lost election to the Presidency of the Royal Society. He had to content himself with emulating Newton in another way, as a Master of the Mint. Herschel made the first photographs on glass in 1839 and in turn was immortalized by sitting as an old man for Julia Cameron.

1742 (2.5 centenary)

Did Edmund Halley die in 1742? He would have said yes, because his death date, 14 January 1743 Gregorian, was 4 January 1742 *stylo anglico*; and we think that his opinion should prevail. Halley contributed largely to the geography and astronomy of his day, not least by persuading Newton to undertake the work of composing the *Principia*. Halley used its principles to argue that the bright comet of 1682 (3.1¢) was identical with those of 1607 and 1531, and probably also with those of 1456, 1380 and 1305. He reckoned, rightly, that it would return in 1758, and regretted that, whether he died in 1742 or 1743, he would not see it happen.

Anders Celsius introduced in 1742 the inverse of the temperature scale that now bears his name. He probably took zero for the hot point and 100 for the cold point in analogy to the Earth's climate, which is torrid at 0° latitude and frigid at 90°. This rational scheme was soon overturned; with the help of the decimalized metric measure, the inverted Celsius system drove its rivals, the scales of Réaumur and Fahrenheit, out of Europe.

Another important pure-and-applied invention of 1742 was the ballistic pendulum as described in Benjamin Robins' *New Principles of Gunnery*. The pendulum, a heavy block of wood hanging from a cord, recoiled when struck to a height that measured the muzzle velocity of the bullet that hit it. Robins investigated small shot; his successors at the Royal Military Academy at Woolwich, particularly Charles Hutton, studied cannon balls and tried to determine the effects of air resistance.

Ballistics was a means of upward mobility for mathematicians. Christian Goldbach entered the Petersburg Academy of Sciences in part on the strength of a recommendation from a General impressed by his knowledge of the subject. The post gave Goldbach occasion to take up such practical problems as Fermat's last theorem. In 1742 he enumerated another, still unproven theorem: every even number can be expressed as the sum of two primes.

While Goldbach was conjecturing comfortably in St Petersburg, explorations commissioned by his academy were exploring Siberia. In 1742, a Navy lieutenant, T. Chelyuskin, rounded the cape now named for him, the most northerly point of the Eurasian landmass, at latitude 77° 43', corresponding to 85 per cent of the way up Celsius' temperature scale.

1692 (3.0 centenary)

This year three centuries ago there died Elias Ashmole, a man perfectly representative of the transition from Renaissance to early-modern science. He began his literary career by translating astrological and alchemical texts. A great amasser of antiquities and natural curiosities, he recognized the importance of permanent and accessible collections for the advance of scientific knowledge; his own collection, given to the University of Oxford, became a centre of study for many who did not share his beliefs in the occult sciences. He saw no contradiction between the then new enthusiasm for experimental natural philosophy and his own interests. He became a founding

fellow of the Royal Society of London and a promoter of its work while labouring to restore astrology to its ancient lustre.

1642 (3.5 centenary)

In keeping with our decision about Halley, we must assign Newton's birthday to 1641, and apologize for not noticing it last year; we realize, however, that our scruples are not likely to weigh much with people bent on a celebration irrespective of anniversarial subtleties. The year belongs to Galileo, who died in Arcetri near Florence, in the house to

Blaise Pascal (died 1662, 3.3¢) started on the less dangerous project of a calculating machine. Three years later, in 1645, he succeeded in designing, manufacturing and distributing a mechanical calculator. Immediately thereafter he took up the problem of the suspension of mercury in an inverted tube, an experiment invented by Galileo's disciple Evangelista Torricelli.

We would be remiss if we did not report that in 1642 Abel Janszoon Tasman discovered the island named after him, and also the west coast of New Zealand. The age of exploration also permitted the Dutch physician Jacob Bontius (born 1592, 4.0¢) to observe and describe beri-beri, and the Spaniard Pedro Barbo to alert the Old World that the bark of cinchona, now known to contain quinine, was useful in the treatment of intermittent fevers.

1592 (4.0 centenary)

In 1592, Galileo made a decisive and symbolic move from a professorship of mathematics at the parochial University of Pisa to a similar position in Padua, the university town of cosmopolitan Venice. The consequent shift in his approach to physical problems may be symbolized by taking 1592 as the weighted-average date of his early work on mechanics. At Pisa, he composed several tracts based on, or critical of, Aristototele's ideas about motion as taught by the Jesuits at the Collegio Romano. In the last of these tracts, written in 1591, he claimed that he had proved by experiment the false proposition that in free fall lighter bodies initially outpace heavier ones. A well-meaning disciple much later corrected the claim and

invented the tall story of the demonstration at the leaning tower of Pisa. In 1593, Galileo began a treatise on simple machines, signalling a shift of interest that would lead him to establish a workshop, design instruments, and do experiments.

Further south, the remains of Pompeii were uncovered when the Italian architect and engineer, Domenico Fontana (born 1543, 4.5¢ ± 0.0270), began tunnelling under a hill.

1542 (4.5 centenary)

This year we can celebrate the teaching as well as the doing of science if we accept the old date for the first publication of Robert Recorde's *The Ground of Arts*. (Modern bibliographers prefer 1543 to 1542.) The *Ground*, an introduction to arithmetic, went through more editions during the next century than even those who have mastered its princi-

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Herschel immortalized by Julia Cameron.

which he had been confined for a decade by decree of the Holy Office. The 350th anniversary of his death is an occasion to reaffirm the right and power of untrammelled thought and to recognize that ridiculing your opponents is not a good way to disseminate your insights. Among Galileo's many achievements we call attention to his analyses of pendular motion, free fall and descent along inclined planes (beginning around 1602, 3.9¢); his brilliant attack on Aristototele, and praise of Archimedes, in the *Discorso*. . . *Intorno Alle Cose che Stanno in sul l'Acqua* (1612, 3.8¢); his mordant and ill-advised diatribe against Jesuit astronomy, *Il saggittatore* (1623, 3.7¢ to 0.3%); and the book that simultaneously ruined and exalted him, the *Diagolo*. . . *sopra i due Massimi Sistemi del Mondo, Tolemaico e Copernicano* (1632, 3.6¢).

Galileo miscalculated the effect of his *Dialogue*. A few months after his death,

ples can count. Like Galileo, Recorde wrote in dialogues to make his wares attractive; and he led his students step by step, beyond fractions, to the elements of the *Elements* of Euclid (*The Pathway to Knowledge*, 1551), the geometry of the sphere (*The Castle of Knowledge*, 1556) and quadratic equations (*The Whetstone of Witte*, 1557). We leave it to our readers to supply the content of another of Recorde's primers, *The Urinal of Physick* (1547).

We approach Columbus, backwards, through two of his successors, Hernando de Soto, born around 1492 (5.0¢), and Fernão Mendes Pinto. In 1532 (4.6¢), de Soto led 300 volunteers to reinforce Pizarro in Peru; he then played a prominent part in the destruction and looting of the kingdom of the Incas. The fortune he made in this work did not satisfy his greed. His quest for the reputed gold of Florida ended not in the money houses of Seville but on the banks of the Mississippi, where he died in 1542. The same year Pinto, an adventurer of an entirely different stamp, reached Japan. He became the friend and travelling companion of St Francis Xavier, enlisted briefly in the Order of Jesus himself, and gave a substantial part of his fortune to the evangelization of Japan. He had accumulated his capital as a merchant and soldier between bouts of slavery, about seventeen in his reckoning. He returned home to marry and to write up his adventures as a reading primer for his children. The result, entitled *Peregrination*, became an international best-seller when published in 1614 by a charitable house for penitent women, to which he had left the manuscript.

Visual evidence that the New World was affecting the Old can be seen in Leonhart Fuchs's *De Historia Stirpium*, due now for 4.5¢ celebration. It brought the art of botanical woodcut illustration to a new high, demonstrated that a picture is really worth a thousand words (the text was as antiquated as the pictures were innovative), and presented visually for the first time Indian corn and the pumpkin.

1492 (5.0 centenary)

Columbus thought he had landed in 1492 where Pinto reached 50 years later. He persisted in this view after he had obtained by observation of eclipses an estimate of the longitude of his landfall in the 'Indies'. On this theory, the Eurasian landmass would end somewhere east of California. As an explorer and exploiter, Columbus resembled Pinto more than de Soto. Although much concerned with titles, fame and wealth, and with the propagation of the Christian faith, his compulsion was to conquer the sea, not the 'Indians'.

Trips across the Ocean Sea required

Ernest Lawrence asking cyclotron builders not to form a union.

better methods of navigation than coastal sailing did. Every explorer needed, and in turn made necessary, a new map, and even new ways of drawing maps. The Portuguese mathematician Pedro Nuñez (born 1502, 4.9¢) observed that a ship sailing at constant bearing (or, what is the same, on a fixed rhumb line) would not describe a great circle on the Earth's surface even if the compass always pointed to geographical north. To find out where a constant rhumb will take you is not an easy task. To simplify it for ships' officers, Gerard Mercator (born 1512, 4.8¢) introduced the cylindrical projection named after him, which represents rhumb lines on a sphere as straight lines on a chart.

The finding of the longitude at sea remained a problem. Galileo thought that his discovery of the Medici stars — the four brightest moons of Jupiter — offered a solution. The times of disappearance and reappearance of the moons as calculated for a given place could be compared with the corresponding times as observed on shipboard; the difference between the calculated and observed times measures the difference in longitude between the places. Galileo's solution failed for lack of sufficiently precise tables. Longitude remained guesswork until the perfection of the marine chronometer and lunar tables around 1772 (2.2¢).

1942 (0.5 centenary)

Among the vital statistics of 1942 were the deaths of Franz Boas, historical-physical anthropologist, and specialist on the Indians of North America; William Henry Bragg, pioneer in X-ray crystallography and British scientific statesman; William Matthew Flinders Petrie, specialist in Egyptian archaeology; and Jean Perrin, who demonstrated the atomic theory of matter; and the birth of cosmologist Stephen William Hawking.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Japanese and US physicists made preparations, not widely known in their day, for events that exhibited the vast range of interactions made possible by their science. For one, Shoichi Sakata and Toshi Inoue suggested that the meson identified in cosmic rays differed from the meson postulated in Hideki Yukawa's theory of nuclear force. They anticipated by five years the development of a two-meson theory in the West, by Robert Marshak and others, and the experimental detection of the mesons (now called, respectively, the μ and the π) by Cecil Powell's group at the University of Bristol.

During 1942, Enrico Fermi, a refugee from fascist Italy, led a team that assembled a pile of 400 tons of graphite, 6 tons of uranium metal, and 50 tons of uranium oxide under the stands of an athletic field at the University of Chicago. The object was to demonstrate the feasibility of a sustained chain reaction that would generate heat and turn the non-fissile heavy plentiful isotope of uranium into the nuclear explosive plutonium. Fifty-one days after Columbus Day, 1942, Fermi carefully started up the reaction by gingerly removing its control rods. After many hours, a sustained reaction set in, which Fermi shut off when it reached a power of one watt because of the intensity of the associated radioactivity. Arthur Compton, the director of the project, ran from the successful test to the telephone. He called the head of the committee overseeing the American uranium project. His announcement made a chilling reference to our anniversary of the year: "The Italian navigator has just landed in the New World." □

J. L. Heilbron is the vice chancellor of the University of California at Berkeley; W. F. Bynum is at The Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.

What promises are there for 1992?

Wishing for better things is an old habit at this time of the year, but wishing for more money is less worthy (and less likely to be worthwhile) than wishing for tangible success.

In the British tradition, a new year is either an occasion for resolutions that would change personal behaviour (as in "I will henceforth deal courteously even with the most importunate authors") or for unfalsifiable prediction (as in "July: great floods; Prince Charles falls from horse; government grapples with new Sterling crisis"). Sadly, in science, resolutions on personal behaviour (such as "I resolve to find the top quark") cannot count for much, while predictions that cannot be falsified have been made disreputable by Karl Popper. May wishes fill a kind of halfway house? What would most improve the temper of the scientific community in 1992?

The chorus "More money!", even when nonsensically corrupted to "More funding!", is readily evoked, but will not suffice. That does not, of course, imply that there is no substance in the complaints from many parts of the research enterprise of shortages of research support. People in biomedical research in the United States, for example, note with sorrow (or, sometimes, anger) that even the budget of the National Institutes of Health cannot provide personal research grants for more than a proportion of the able people who would have won them in earlier times. (Two years ago, Leon Lederman made the same argument on behalf of research in general.) There may be changes in the federal budget due a month or so from now, but they will be marginal. Nor should it be otherwise when President George Bush's eagerness to 'kick-start' the US economy, and to launch a roseate re-election year, must be tempered by the size of the federal deficit.

In Britain, used for longer to the notion that the administration of science is like the cutting of a wedding-cake, only continued habitual gentility suppresses what should now be the general complaint — that the personal impoverishment of researchers (by means of paltry salaries, but in other ways as well) offends against the public doctrine that "a nation that lives by its brains" must nurture them as well. It will be no bad thing for Britain if 1992 sees protests on this score which are less passive than the soft option of westward migration.

In France, by the magic that even modest continued growth can bring, complaints centre more on administration (of the universities, for example) than on money. Italy's continuing economic miracle (perpetual trouble-free deficit financing) compared with which that of the early 1960s

was a flash in the pan, seems also to leave room for doing things for the first time. And while the glumness in some laboratories in western Germany occasioned by the needs of unification with the east will attract some sympathy, most will admire both the ingenuity with which organizations such as the Max-Planck Gesellschaft (traditionally a backer of institutes) have been able to wish good eastern research groups onto still-renascent universities and the Bundesbank's stolid defence of the Deutschmark, the week before Christmas, which seems certain to make eastern Germany as prosperous as anywhere else in Europe within a few years.

Perhaps what the scientific community most needs, in 1992, is a better understanding of the inevitable interaction between the scale of research support and the state of the economy in which it is embedded. There is an inherent weakness in the argument for research, which is a seamless web stretching from higher education to the laboratory bench. Ups and downs in the scale of support, such as those Britain suffered in the early 1980s, destroy the continuity from which productive science flows. Everybody in research knows that.

But because the benefits of research are inevitably long term, governments in a jam (such as that of the new Russian republic) inevitably first think of jettisoning research in the here and now, supposing that they can buy replacements from some shelf when things get better. The danger to the research enterprise is greatest when governments are battling against inflation. In the United States and Britain now, where governments are looking for ways of spending money without rekindling inflation, increasing research support should be at least as deserving a counter-cyclical candidate as more spending on the construction industry (with which the US government is dicker). Keynes would agree, were he alive.

Japan, as always, is a different case. Belatedly, the government has woken up to the need to strengthen research in the universities just as the US government has appreciated the difficulties of building the Superconducting Super Collider within the constraints of the federal deficit. (Whatever happened to that peace-dividend, some will be asking?) Now it seems (see page 8) that Japan may make a substantial contribution to the cost. One wish for 1992 is that SSC should now be built. Another must be that this unseemly epi-

sode should not permanently sour the Japanese community's view of US science.

The case of science in eastern Europe and the former Soviet Union is much more serious. It is inevitable that many people now working as professional researchers will have to turn to other things, either for survival or because their governments (usually their direct employers) have no need of them. This upheaval is being further complicated by preoccupations with the recent past. Does equity require that ex-Party members, once favoured above others, should now run on an outside track? Or even preferentially lose their jobs, merit notwithstanding? It is probably vain to hope that this process will not cause lasting damage.

But there is a chance that something might be done about Europe's research policy, as represented by that of the European Commission (see page 3). Europe's great showdown at Maastricht a month ago seems to have left most things and most policies unchanged. The European Commission seems still wedded to its policy of supporting industrial research involving transnational collaboration even though the single market whose aim is to abolish national frontiers (for all economic and many social purposes) is only a year away. Is it too much to wish that the commission will realize, before too much time has passed, that its chief role should be the support of basic research?

Such wishes may be helped to come true by argument and persuasion, but even if all of them came true, the scientific community would not recover the sense of wellbeing it has enjoyed in even recent decades, the 1970s, for example. That is a more profound reason why mere money (and the perceived shortage thereof) is only part of the story. A few tangible successes would work greater wonders.

If the Hubble Space Telescope had functioned as planned, we might all now be basking in novel contemplation of how the Universe is constructed. Maybe 1992 will settle the issue of the Big Bang, one way or the other. Or questions of gene regulation, already answerable, will be answered simply, to the general enlightenment. All wish-lists will be a little like that, but what would make the greatest difference to the reputation of science and the esteem of its practitioners would be an effective and simple way of treating people with AIDS. Is that too much to ask for?

John Maddox

Warning — handle with care!

Ian Stewart

A NEW area of applied mathematics is emerging from the interaction of dynamical systems theory and numerical analysis. Traditionally, numerical methods are used — among other things — to solve differential equations that arise in dynamics; but in the new interaction the flow is the other way, using dynamics to illuminate how numerical methods work. A new paper by H. C. Yee, P. K. Sweby and D. F. Griffiths (*J. comput. Phys.* **97**, 249–310; 1991) applies the ‘dynamics of numerics’ point of view to several popular numerical differential-equation solvers, with the disturbing conclusion that standard practice may sometimes be seriously misleading, and

that much work is needed to find out why and to stop it happening.

Numerical methods typically start with a continuum model, such as an ordinary or partial differential equation, and turn it into a discrete one in both space and time: a continuous domain such as a rectangle is replaced by a rectangular grid of discrete points, and time is deemed to flow in tiny steps. The value of a given variable is calculated from its values at nearby points and at previous times. Variable or non-uniform grids and variable timesteps are also used. ‘Discretization’ is essential for computer implementation and cannot be dispensed with. The essence of the difficulty is that the

dynamics of discrete systems is only loosely related to that of continuous systems — indeed the dynamics of discrete systems is far richer than that of their continuous counterparts — and the approximations involved can create spurious solutions.

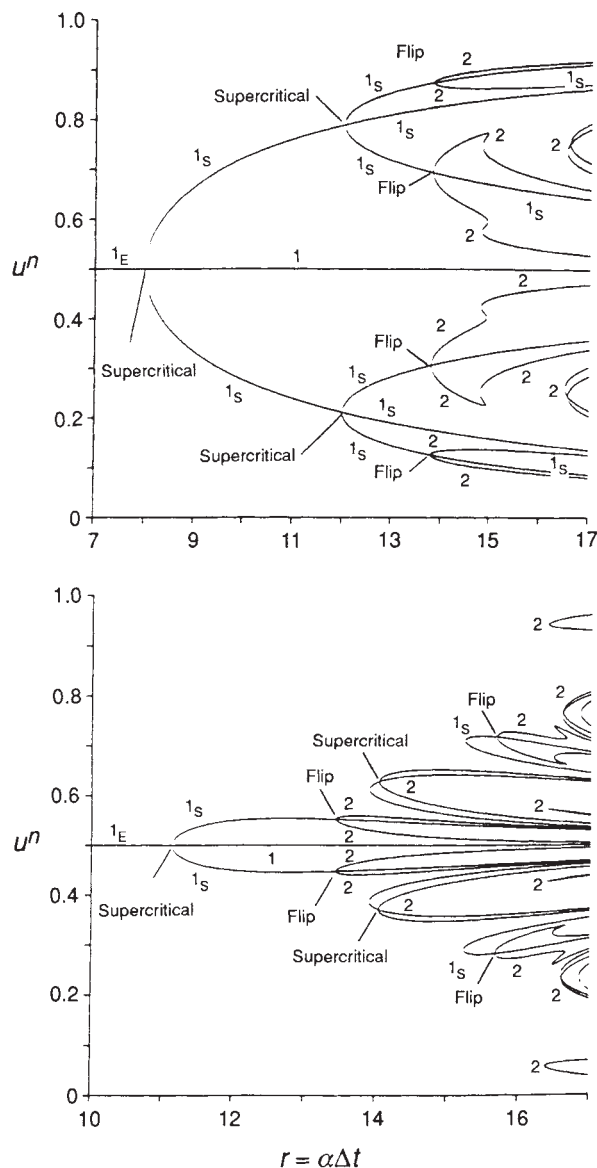
The method used by Yee *et al.* is to take continuum models with known explicit solutions (the most straightforward way of being sure what is ‘really’ happening), discretize them according to various standard numerical methods, and apply techniques from discrete dynamics to analyse the behaviour of the resulting systems. The authors concentrate on ordinary differential equations: future papers will investigate other types of numerical problem. Three standard differential-equation solvers are the Euler, leapfrog, and Adams–Bashforth methods. When applied to the logistic equation $du/dt = u(1-u)$ they correctly compute fixed points (steady states) for the equation but can generate spurious periodic points. The modified Euler, improved Euler, fourth-order Runge–Kutta and predictor–corrector methods introduce spurious steady states. The maximum number of steady states ‘found’ by these methods varies between 4 and 16 depending on the method used. For the

ordinary differential equation $du/dt = au(1-u)(b-u)$ it varies between 9 and 81. The figure compares the results of two methods on this equation, as α varies, when $b = 0.5$.

One objection may be raised to some of these results. It is common practice to use a quantity known as the linearized stability limit to test for possible numerical instability. In many cases the spurious steady states lie beyond this limit, so conventional numerical analysis would anticipate problems. However, both the modified Euler scheme and the fourth order Runge–Kutta scheme can produce spurious steady states *below* the linearized stability limit. Moreover, the effects of operating above the linearized stability limit are not necessarily divergent solutions, as often assumed, but can be convergence to spurious steady states. In short, no obvious warning signs may be seen even though the method is generating nonsense. Although the examples used in the paper are deliberately kept simple to permit explicit analysis, the same problems are more, rather than less, likely in more complicated models.

Other dynamical features, such as basins of attraction (which initial conditions lead to which steady states?) are equally sensitive to the choice of numerical method. Non-unique steady states can be introduced by discretization, even though the original continuum equation has a unique solution. Spurious periodic solutions can be generated by discretization of nonlinear partial differential equations. And if regular dynamics creates such havoc, what of chaos? The increasingly popular technique of using numerical solutions of partial differential equations to generate time series that are then analysed by signal-processing methods to obtain information about the original continuum model (such as which ‘modes’ are operating) may be seriously misleading. Contrary to popular belief, errors of this type may not be cured by reducing the time-step or grid size.

The dangers are especially great in computational fluid dynamics (CFD). The type of problem studied using CFD has changed dramatically over the past few years. But many algorithms widely used in CFD were originally designed for much simpler problems, such as perfect or ideal gas flows. Yee *et al.* remark that “A straightforward application of these numerical methods to nonequilibrium flow or combustion related model problems can lead to wrong results, slow convergence, or even nonconvergent solutions”, and that “new algorithms and/or modification and improvement to existing numerical methods . . . are urgently needed”. Guidelines developed using linearization methods are not always valid for nonlinear problems. Even conventionally well-informed use of con-



Stable and unstable steady states and period-2 points of the equation $du/dt = \alpha u(1-u)(0.5-u)$ according to the improved Euler (top) and fourth-order Runge–Kutta (below) methods.

ventional methods may lead to nonsense on unconventional problems.

Yee *et al.* make a number of general recommendations for avoiding trouble, but conclude that the safest route is to have some understanding of the dynamical behaviour of the numerical method being used. In short, the whole area needs sorting out. The main problems are not so much numerical as dynamical: the actual behaviour of the

continuum models, and their relation to discretizations, must be the central object of study. Until these advances in the dynamics of numerics are made, all users of the numerics of dynamics — most of whom are wielding a mathematical tool without understanding how it works — should heed the warning. □

Ian Stewart is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

CELLULAR BIOPHYSICS

Proteins as machines

Thomas D. Pollard

THIRTEEN different types of protein machine, the driving units that produce macromolecular movement in living organisms, came under scrutiny at the sixth molecular biology symposium at Indiana University late last year*. The upshot is a clearer view of the factors that limit our understanding of such machines, and the identification of a major conceptual issue that will have to be dealt with before progress can be made on a broad front (of which more below).

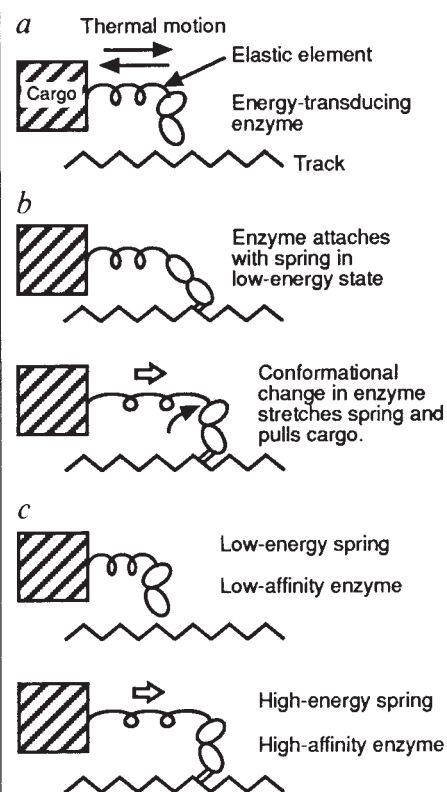
Molecular machines act like engines to pull intracellular cargo along protein polymer tracks, or like processors on an assembly line to copy the base sequence of DNAs into RNA or DNA replicas, or like selective valves or pumps to transfer macromolecules across lipid membranes from one intracellular compartment to another. From the data we have, these machines look like so many apples and oranges; for example their immediate energy sources include ATP, GTP, proton gradients, membrane potentials and perhaps even protein folding. Nevertheless, they share two general features. All of them appeared early in evolution, many in bacteria and others in the first nucleated cells, so that once a basic mechanism is understood it will be widely applicable; and (fortunately) all are robust enough to withstand isolation and reconstitution in the test tube, which allows detailed biochemical characterization of the individual steps taking place. In every case, however, we lack one or more of four types of information necessary to explain the mechanisms operating at the molecular level.

First, we must have an inventory of the parts, which is best achieved by a combination of genetics and biochemistry. A great deal of progress has been made in this respect, although the challenge varies considerably. For example, a small 43K fragment of kinesin can move microtubules (L. Goldstein, Harvard); more than 10 weakly associated proteins contribute to DNA replication

(B. Alberts, University of California San Francisco); and dozens of polypeptides are required to form nuclear pores with a mass of 100,000K (D. Forbes, University of California San Diego).

Second is the identification of the chemical intermediates in the reactions carried out by each machine. Here biochemistry has been the classical approach, but genetics can offer decisive insights when a mutation stops a process in midstream, for instance in mitochondrial protein import (N. Pfanner, University of Munich). In his new work on RNA polymerase, M. Chamberlin (University of California Berkeley) has used biochemical tricks to stall transcription at a specific base on DNA and synchronize subsequent steps for analysis, and a similar approach has been used with the endoplasmic reticulum protein translocator (G. Blobel, Rockefeller University; H. Bernstein, University of California San Francisco), DNA polymerase, myosin and dynein. But information on chemical intermediates remains incomplete in most cases.

Third, we must measure the rate constants for the transitions between each of the intermediates. This requires chemical kinetics, an approach with an undue reputation for difficulty among biologists, yet rate constants are nonetheless essential to understand the pathways of a reaction through the possible chemical intermediates. As a bonus, the ratio of the rate constants for each step gives the equilibrium constant, and thus the free energy change, by which the steps where energy is being consumed or work may be done can be identified. New results with RecBCD (G. Smith, Hutchinson Cancer Center) and nuclear pores (Forbes) provide examples where decisive observations could be obtained by transient kinetic analysis of a single step. Even when this information is available, however, its interpretation may be challenging. Thus most of the kinetic data on myosin come from proteins in solution, and further data are needed to learn how the mechanical forces that exist when the



Schematic diagrams of protein machines. *a*, The energy-transducing enzyme diffuses near its track while attached to its cargo by an elastic element. *b*, A conformational model for energy transduction. The enzyme attaches to its track with its spring in a low-energy state and then uses its energy source to drive a conformational change that stretches the spring and pulls on the cargo. *c*, A biased attachment model. The energy source is used to make a subtle change in the enzyme such that its probability of attachment to the track is highest when the spring is stretched. (More detailed treatment of these alternatives is given by A. F. Huxley and R. M. Simmons, *Cold Spring Harbor Symp. quant. Biol.* **37**, 669–680; 1972.)

machine is operating might influence the kinetic constants.

Finally, the detailed structure of the machine itself is needed to appreciate how the chemical reactions are coupled to work. This, above all, is the limiting factor on our understanding — atomic resolution is required, and only X-ray crystallography and NMR spectroscopy are equal to the task. An atomic model of the actin filament (K. Holmes, Max Planck Institute, Heidelberg) has provided ideas about how myosin might move along this molecular track. But without knowing the structure of the enzymatic motor myosin, it is impossible to relate the steps in the hydrolysis of ATP to the mechanical output of the system, even with *in vitro* assays for the force (T. Yanagida, Osaka University) and motion (J. Spudich, Stanford University) produced by individual molecules during each round of ATP hydrolysis (this topic was discussed recently

* *Proteins as Machines*, Indiana University, 27–30 October 1991.

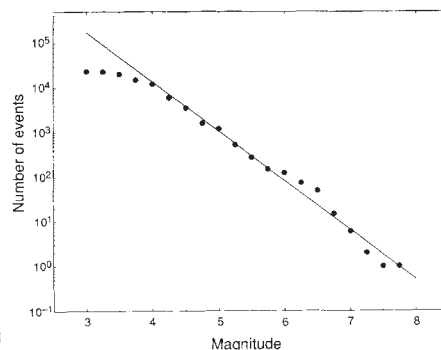
in News and Views by M. Irving, *Nature* **352**, 284–286; 1991). Insights from atomic structure on the binding of RecA to DNA (T. Steitz, Yale University; *Nature*, in the press) indicate the direction that others will have to take.

The discussion of myosin and kinesin (R. Vale, University of California San Francisco) highlighted the major conceptual question of the conference, which was posed by H. Berg (Harvard University). Do the protein machines use their source of chemical energy to produce movement purely by internal changes of shape (a conformational mechanism), or is the energy used to bias the attachment of the components of the machine to those states where random thermal motions have put the machine into a high energy state? (See figure.)

In the case of the rotary motor that powers the bacterial flagella, Berg

argued in favour of the second process, but the issue in particular and in general is far from settled because there is no well-documented model for either mechanism. Yanagida argues that the motion produced by myosin far exceeds the distance that could be achieved by a conformational change, whereas most biophysicists are more comfortable with conformational changes such as the rigid body rotations of 10° to 18° that occur between two domains of enzymes (hexokinase, aspartate aminotransferase) during their interaction with substrates. Stay tuned. This is one of the big unanswered questions in biology and biophysics. □

Thomas D. Pollard is in the Department of Cell Biology and Anatomy, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA.



Schematic frequency-magnitude plot, based on artificial but realistic earthquake counts. The low-magnitude roll-off occurs at magnitude 4, and we assume that the observing network detected only some of the events smaller than this. The data are scattered at the higher magnitudes, in part because the magnitudes are uncertain and in part because the time interval over which the data were collected is too short to give accurate counts of the less frequent larger events. The calculated *b*-value is 1.11. Magnitude 7.75 is the largest event in this data set, but the extrapolation of the linear fit calls for half a magnitude 8 event in the time interval, or one magnitude 8, on average, in twice the time covered by the data set. This extrapolation is uncertain because it is not known from the data if a magnitude 8 will ever occur, and the work of Pacheco *et al.* shows that even if an 8 is to be expected on other grounds, the average inter-event time based on extrapolation is invalid.

EARTHQUAKES

Sizing up the threat

Carl Kisslinger

How does the number of small earthquakes that occur compare with the number of big ones? Even a cursory examination of earthquake counts reveals that the small ones are much more frequent. But Pacheco, Scholz and Sykes show on page 71 of this issue¹ that it is not possible simply to extrapolate from the frequency of the smaller earthquakes to determine the threat of the rarer large ones.

The distributions of earthquakes in terms of energy release, as well as in space and time, provide a basis for both physical understanding of earthquake processes and efforts to assess the likelihood of occurrence of future strong events. One of the accepted characteristics of earthquakes is that when $N(M)$, the number of earthquakes with magnitude equal to or exceeding a value, M , is plotted against the magnitude, a simple relation emerges. Known in the West as the Gutenberg–Richter relation, and in Japan as the Ishimoto–Iida relation, this is written as $\log N(M) = A - bM$, where b (the ‘*b*-value’) is a number close to 1 and A describes the level of activity.

This relation is used to describe earthquake distributions globally or in limited regions, large and small, with values of A and b determined from the corresponding data. It is the danger of pushing this relation too far that Pacheco *et al.* point out in this issue. They demonstrate from analysis of reliable earthquake catalogues, including their own recently revised listing of strong earthquakes², that the conventional assumption that the *b*-value is a constant over a wide

range of magnitudes, from the smallest earthquakes observed in the selected area to the largest, is not valid. Furthermore, they provide a physically sound reason for why it should not be.

When $\log N$ is plotted against magnitude (see figure), the data tend to follow a linear trend over some part of the range of magnitudes. But with an excessive magnitude range, the data often fall below this line for small events, and do not fit it very well at the top end, having too many or too few of the largest events. The reasons offered for these departures are incompleteness of the catalogue of earthquakes used and problems with the way magnitudes are measured. Seismic networks may fail to catch some of the smaller events, and the catalogue may cover too short a period to yield an accurate count of the largest earthquakes. Further, the magnitude scales in common use cannot cope with the most energetic events, so that the calculated magnitude does not increase as the energy released does. It has become almost a matter of faith with many seismologists that a perfect catalogue, with correctly determined magnitudes, covering all time would yield data that fit the line from end to end.

This frequency-magnitude relationship has important applications and implications for seismology. One is application to the problem of estimating the probability of occurrence of future events with magnitudes exceeding some selected value. If the counts are normalized to unit time, say one year, then the average time between events at any

magnitude level can be taken directly from the graph. The mean time between events, determined this way or by other approaches, is essential input to calculations of the conditional probability of earthquake occurrence³.

For many seismically active regions, available data include adequate numbers of only moderate to strong events. There is a strong temptation to extrapolate the graph to the great earthquakes, often the ones of most interest in prediction. If, as Pacheco *et al.* show, a single *b*-value does not characterize the entire magnitude range, the lapse time derived by the extension of the fit to moderate earthquakes and any probability based on it are wrong.

Also confronting earthquake hazard estimation at the high end of the magnitude scale, is the problem of the third, implicit parameter that fixes the frequency-magnitude curve: the maximum magnitude that will ever occur within the region under study. Although nothing in the definition of the magnitude scale sets an upper limit, nature seems to have done so, somewhere around magnitude 9 for the whole Earth. Maximum probable earthquakes in specific regions are estimated from the statistics based on the history of the site in question, but a deterministic approach based on examination of the geological structures that will produce the earth-

ECOLOGY

quakes has much to recommend it.

Numerous studies have analysed b -values as an indicator of physical conditions in a seismic zone. High b -values indicate a fault zone that is highly fractured, with variable properties along its length⁴. Regional differences of the b -value have been related to differences in the ambient shear stress⁵. Some investigations have found a significant difference in the regional b -value for events preceding and following a great earthquake⁶. It is not likely that the result of Pacheco *et al.* will influence the application of the frequency-magnitude relation to problems of this sort but future investigators may be more careful in how they pick the magnitude limits when calculating b -values.

In the search for a fundamental understanding of earthquakes, the notion of a constant b -value over a wide range of magnitudes has been used to argue the self-similarity and fractal nature of earthquake sources⁷⁻⁹. An implication is that we can learn the essential physics of great earthquakes by studying the much more frequent small ones, even down to the scale of rock-fracturing experiments in the laboratory. As Pacheco *et al.* argue, similarity between small and large events breaks down when the width of the earthquake rupture extends through the entire seismogenic zone, so that the width is bounded but the length of the rupture can increase. A change in b -value for large events has been suspected previously, but the improved data, including improved magnitudes, in this work strongly support the conclusion.

The break in slope for the frequency-magnitude curve at high magnitudes requires a re-examination of both procedures and conclusions developed empirically in seismology on the basis of the classical model. On the other hand, the large number of b -value analyses that have been done with maximum magnitudes up to 6 or 6.5 remain valid as long as one does not extrapolate to events much larger than those in the data set. □

Carl Kisslinger is in the Department of Geological Sciences and the Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Colorado 80309, USA.

Feeble links in food webs

John H. Lawton

THE meek may inherit the Earth, but the weak already appear to have a working majority. On page 73 of this issue¹, R. T. Paine provides estimates for the effects of individual herbivores of seven species on populations of their algal food resources. Mostly the effects are feeble. Although there have been numerous experimental studies investigating the influence of individual enemy species on one or more victim species² (the principal actor in a classical study, again carried out by Paine³, is shown in the figure), nobody has previously examined per capita effects for so large a set of

more strongly species interact (the larger the per capita effect of species A on species B), the more difficult it is to find stable, persistent combinations of species.

In the absence of any better information, dynamic models have usually drawn their estimates of per capita interaction terms from a uniform or at least a symmetrical distribution, in which only a small proportion of terms are zero or close to zero. Paine now shows that, at least for the algal grazers in the intertidal zone of Tatoosh Island, Washington State, the usual assumptions about interaction strengths and symmetry are false.

The average per capita influence of all the grazers on recruitment by brown algae was weakly negative, but close to zero. A net negative value should, of course, characterize the impact of a guild of consumers on their food resources; but few food-web modellers have imagined average interaction strengths to be so low. More importantly, estimates for individual species were highly skewed round this average value. Five out of the seven species (two chitons and three limpets) were relatively benign, and only two species (the urchin *Strongylocentrotus purpuratus* and the chiton *Katharina tunicata*) had strong negative effects. Interestingly, some of the herbivores had weak positive effects on populations of brown algae because, by grazing competing crustose coralline algae, recruitment of brown algae was in-



Star performer — *Pisaster ochraceus* is an example of a predator species that has a considerable effect on its prey³; such strong interactions may involve only a minority of consumers. (Photograph by J. H. Lawton.)

co-existing consumers sharing a common food resource.

It is important to know the strength of species' interactions, because one group of theoretical models that predicts patterns in food webs⁴ has the per capita impact of consumers on consumed populations as a critical parameter. Predictions about the structure of food webs are sensitive to the magnitude of these interaction terms^{5,6}. A closely related family of models that explores relationships between species richness and community dynamics (the diversity:stability problem for example⁷) also makes assumptions about the strength of species' interactions. In general terms, the

increased. Such counter-intuitive responses have been increasingly observed and predicted⁸ in field experiments, driven by indirect interactions between members of the food web. They can make the interpretation of 'simple' field experiments extremely difficult⁸.

Two other lessons emerge from Paine's study. One is that the strong per capita effects of *Strongylocentrotus* and *Katharina* are not additive, because the urchin competitively excludes the chiton at normal densities. The second is that interactions are not uniformly distributed in space; the effects of *Strongylocentrotus*, in particular, are strongly aggregated. Food-web modellers

1. Pacheco, J. F., Sholz, C. H. & Sykes, L. R. *Nature* **355**, 71-73 (1992).
2. Pacheco, J. F. & Sykes, L. R. *Bull. seism. Soc. Am.* (submitted).
3. Nishenko, S. P. & Buland, R. *Bull. seism. Soc. Am.* **77**, 1382-1399 (1987).
4. Mogi, K. *Bull. Earthq. Res. Inst., Tokyo Univ.* **40**, 125-173 (1962).
5. Wyss, M. *Geophys. J. R. astr. Soc.* **31**, 341-359 (1973).
6. Suyehiro, S., Asada, T. & Ohtake, M. *Pap. Meteorol. Geophys.* **15**, 71-88 (1964).
7. Aki, K. in *Earthquake Prediction, An International Review*, Maurice Ewing Series 4, American Geophysical Union, 566-574 (1981).
8. King, G. *Pure appl. Geophys.* **121**, 761-815 (1983).
9. Scholz, C. H. *The Mechanics of Earthquakes* 187-189 (Cambridge University Press, 1990).

have not explored the consequences of aggregated attacks, but incorporating aggregation into them will almost certainly result in a change in some of the model predictions⁶.

Although Paine is cautious not to generalize these results too far, it seems unlikely that intertidal grazers are wildly atypical. The literature on the effects of individual species² suggests that most enemies have a significant influence on victim populations; unfortunately these one-off studies are probably biased by the deliberate selection of consumers likely to yield 'interesting' results during the lifetime of a research grant or doctoral thesis. It is rare for scientists to study non-events; no news may be good news, but it doesn't sell papers. In reality, most feeding interactions are probably feeble. If so, then many of the patterns recently documented in food webs⁴ may have non-dynamic explanations, for example as predicted by Cohen's cascade model⁹. The cascade model generates many of the patterns observed in food webs by making the simple assumption that feeding links are drawn at random (to a given level of web-connectance) from a hierarchy in which species can feed on any species below them in the hierarchy, and are, in

turn, fed on by any species above them. Rare, strong dynamic interactions may serve mainly to reinforce patterns generated by the static cascade model^{4,6}.

Ecology is sometimes criticized for a lack of communication between theoreticians and empiricists. Paine's work is a happy counter-example in which theory has inspired field experiments, which in turn point to new directions for theoreticians to explore. In nature, most interactions may be weak and feeble, but in this area, at least, the interaction between theory and experiment is strong and healthy. □

John H. Lawton is at the NERC Centre for Population Biology, Department of Biology, Imperial College, Silwood Park, Ascot, Berkshire SL5 7PY, UK.

1. Paine, R. T. *Nature* **355**, 73–75 (1992).
2. Sih, A., Crowley, P., McPeck, M., Petranka, J. & Strohmeier, K. A. *Rev. Ecol. Syst.* **16**, 269–311 (1985).
3. Paine, R. T. *Am. Nat.* **100**, 65–75 (1966).
4. Pimm, S. L., Lawton, J. H. & Cohen, J. E. *Nature* **350**, 669–674 (1991).
5. Pimm, S. L. *Food Webs* (Chapman & Hall, London, 1982).
6. Lawton, J. H. in *Ecological Concepts* (ed. Cherrett, J. M.) 43–78 (Blackwell Scientific, Oxford, 1989).
7. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton University Press, 1973).
8. Yodanis, P. *Ecology* **69**, 508–515 (1988).
9. Cohen, J. E., Briand, F. & Newman, C. M. *Community Food Webs: Data and Theory* (Springer, Berlin, 1990).

METEORITES

Stardust and planet dust

T. A. McGlynn

THE constant rain of small particles that bombards the Earth's upper atmosphere may give astronomers a novel method of looking for other solar systems in our Galaxy. In *Icarus*¹, Thomas J. Brophy examines evidence that some of these particles are debris from the formation of planetary systems around other stars. Meteoroid impacts on the Long Duration Exposure Facility (LDEF) satellite and radar tracking of meteoroids indicate that some interplanetary dust travels at speeds greater than the escape velocity of the Solar System. Because these meteoroids with 'hyperbolic' orbits are larger than interstellar grains are expected to be, it is tempting to suppose that they come from other solar systems.

According to current theories, planets grew in the young Solar System from small planetesimals which gradually agglomerated. Many planetesimals were ejected from the Solar System in gravitational encounters with the forming planets. Some which did not completely escape populated the 'Oort' cloud that surrounds the Solar System and is the source of long-period comets. So long as the planetesimal was much smaller than the planet it encountered, the dynamics of the interaction were unaffected by the

planetesimal mass, so that particles from a few micrometres to many kilometres in diameter would have been ejected.

If planetary formation is common in our Galaxy, there should be a significant interstellar flux of this ejected debris from other solar systems. But astronomers have searched in vain² for the larger ejecta which should appear occasionally as comets passing through our own Solar System with velocities in excess of the escape velocity from the Sun.

Brophy shows that for reasonable models of the distribution of sizes, the vast flux of tiny particles expected from other solar systems should more than compensate for their relative invisibility: looking at meteoroids and small meteorites may be a more effective way of detecting extrasolar matter passing through the Solar System than seeking comets on hyperbolic orbits. One of the interesting implications of this work is that planetary formation may significantly contribute to interstellar dust, contrary to earlier pronouncements³. This is especially true for particles more than 1 µm across.

The LDEF observations are of particles of approximately 5–10 µm diameter, and the radar observations are of parti-



Time exposure photograph of a spectacular meteor shower. Could some meteors come from other solar systems?

cles of the order of 0.1 mm. A simple power law extrapolation of the fluxes measured at these sizes to meteorites implies that a significant fraction of meteorites collected may come from other solar systems.

But the simplest models for the detections are not entirely satisfactory. First, if the flux does obey a single power law, an extrapolation from the micrometeorite flux would indicate that each forming planetary system ejects the equivalent of 640 Earth masses of debris, more than is comfortable for current models of the formation of such systems². Second, the extrapolated flux of extrasolar comets should have permitted at least some detections. And third, the mechanism that expelled such particles from other solar systems would also have populated the Oort cloud with similar micrometeoroids which we would readily recognize returning with the solar escape velocity.

Brophy suggests a way out of this contradiction. The planetesimals that populate the Oort cloud and its equivalents around other stars, and which would be seen as extrasolar comets, originate primarily in the peripheries of forming solar systems. Brophy suggests that collisionally evolved grains, perhaps equivalent to a few Earth masses, could come from the counterpart of our asteroid belt at the border between the inner and outer planetary system. This would increase the number of small grains. The large flux of extrasolar meteorites would then not imply an excessive total flux of material from other planetary systems and would not conflict with the lack of extrasolar comets.

There are serious difficulties in extrapolating the measured rates of high-

velocity meteoroids near the Earth back to the flux of material ejected by other solar systems. Small particles will interact with the interstellar medium (ISM) and solar wind and may have their physical and chemical characteristics significantly changed. Even if the particles survive the hazards of the ISM, drag will significantly alter their dynamics⁴. Also, although grains formed in the ISM or in stellar atmospheres are thought to be smaller than 1 μm , a small population of larger native interstellar grains could contaminate Brophy's results. Brophy recognizes these problems, but dealing with them quantitatively is beyond the scope of his paper, so that current results must be considered very preliminary, particularly in any elaboration beyond the simplest models. A comprehensive study of the evolution of these particles in the interstellar medium is needed.

On the observational side too, some scepticism is justified. Current observations of hyperbolic velocities are not without problems; for example, it is difficult to distinguish hyperbolic orbits from retrograde ones, for which the velocity of the Earth is added to the velocity of the particle. Neither of the studies on which this work is based^{5,6} has yet been published in the refereed literature. More independent investigations of the meteoroid flux are needed to identify extrasolar interlopers unambiguously.

Notwithstanding these problems, this approach is a necessary complement to the standard methods of looking for other solar systems — looking for periodic variations in the proper motion⁷ and radial velocity⁸ of stars, or seeking faint companions⁹. Although detection of extrasolar meteorites will not tell us specifically where other planetary systems are to be found, their chemical and physical composition may confirm that the processes which led to our Solar System are commonplace throughout the Galaxy. Extrasolar meteorites (and just possibly comets) will be the only way we will get samples from other solar systems for the next century or more. □

T. A. McGlynn is with Computer Sciences Corporation at the NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

Miniatures get bigger

Frances A. Edwards

LONG-term potentiation of synaptic transmission in the vertebrate central nervous system (LTP), a phenomenon loosely implicated in learning and memory, remains the centre of dispute despite all the attention accorded it over the past couple of years. The dominant question is whether LTP in the hippocampal CA1 region of the brain is due to a change in release of glutamate neurotransmitter from the presynaptic terminal, or to a change in postsynaptic response. On page 50 of this issue¹, Manabe, Renner and Nicoll present the strongest evidence so far that a major factor in the maintenance of LTP is a change in postsynaptic response.

Over the past 18 months several studies have employed classical quantal analysis of evoked currents or potentials in attempts to differentiate between pre- and postsynaptic maintenance of LTP²⁻⁴. However, as previously discussed here⁵ and elsewhere⁶, the interpretation of such results depends on assumptions as to how a synapse in the central nervous system functions. Although the small quantal size observed at such synapses implies that central and peripheral synapses must operate differently, the exact mechanism and consequently the interpretation of quantal parameters remain controversial. Moreover, the details of the analyses themselves have been questioned⁷ and the results and conclusions drawn from different studies have been contradictory.

Manabe *et al.* have avoided quantal analysis and instead have concentrated on a more easily interpreted measurement, the amplitude of miniature postsynaptic currents (miniatures). For all the mechanisms so far hypothesized for central synaptic transmission, a change in miniature amplitude would be interpreted as a postsynaptic effect — either as a change in the number of postsynaptic receptor channels activated by release of a single vesicle or as a change in the single-channel conductance.

Miniatures have not been used before because direct study of them is incompatible with the induction of LTP. They can be measured only when generation of action potentials in the presynaptic neuron is blocked, for example with tetrodotoxin (TTX). As the generation of LTP is dependent on presynaptic activity, this presents a considerable obstacle.

Although it is impossible to avoid the problem completely, Manabe *et al.* have largely managed to circumvent it by using several parallel approaches all approximating measurement of minia-

tures before and after induction of standard LTP. Their approach is ingenious and, to my mind, convincing.

Perhaps the most persuasive finding comes from the comparison of spontaneously occurring events (miniatures and events due to spontaneous presynaptic action potentials) and true miniatures (with TTX). Under control conditions, addition of TTX would be expected to block the larger currents in the spontaneous amplitude distribution. The demonstration that the largest miniature currents (with TTX) occurring after induction of LTP are larger than the largest spontaneously occurring currents (without TTX) before potentiation is clear. This result is further strengthened by the observation of an increase in miniature amplitude induced by bath application of NMDA. Together I think the results leave little doubt that LTP in the CA1 region of the hippocampus results in an increase in the amplitude of miniature currents and therefore, most probably, in an increase in postsynaptic sensitivity.

Would it be possible to interpret an increase in miniature amplitude as a presynaptic effect? The answer, of course, is 'yes'. All sorts of mechanisms for synaptic transmission can be proposed. For example, the skew of the amplitude distribution of miniatures, which is generally observed in central synapses, could be explained by the simultaneous release of several vesicles. Increasing the probability of multiple release would change the average miniature size, presynaptically. Moreover, if the small quantal size were determined by a small and precise number of glutamate molecules in the presynaptic vesicle, increasing this number would also increase the miniature amplitude by way of a presynaptic mechanism.

However, such interpretations have not yet been necessary to explain the data. Whether taking either the more traditional view of a presynaptic determination of quantal size, which has not been disproved at this synapse, or the more recent postsynaptic saturation hypothesis (for a summary of both these hypotheses see ref. 4), the usual

1. Brophy, T. J. *Icarus* **94**, 250–254 (1991).
2. Weissman, P. R. *Nature* **344**, 825–830 (1990).
3. Stern, A. & Shull, J. M. *Astrophys. J.* **359**, 506–511 (1990).
4. Stern, A. *Icarus* **84**, 447–466 (1990).
5. Stohl, J. in *Comets and Meteors II* (eds Lagerkvist, C. I. *et al.*) 565–574 (Astronomiska Observatoriet, Uppsala, 1986).
6. McDonnell, J. A. M. in *IAU Colloq. 126: Origins and Evolution of Interplanetary Dust* (in the press).
7. Harrington, R. S. in *Astrophysics of Brown Dwarfs* (eds Kafatos, D. *et al.*) 3–8 (Cambridge University Press, 1986).
8. Marcy, G. W. & Benitz, K. J. *Astrophys. J.* **344**, 441–453 (1989).
9. Zuckerman, B. & Beckler, E. E. *Astrophys. J.* **319**, L99–L102 (1987).

1. Manabe, T., Renner, P. & Nicoll, R. A. *Nature* **355**, 50–55 (1992).
2. Bekkers, J. M. & Stevens, C. F. *Nature* **346**, 724–729 (1990).
3. Malinow, R. & Tsien, R. W. *Nature* **346**, 177–180 (1990).
4. Foster, T. C. & McNaughton, B. L. *Hippocampus* **1**, 79–91 (1991).
5. Edwards, F. A. *Nature* **350**, 271–272 (1991).
6. Korn, H. & Faber, D. *Trends Neurosci.* **14**, 439–445 (1991).
7. Larkman, A., Stratford, K. & Jack, J. *Nature* **350**, 344–347 (1991).

Birth and HIV

J. J. GOEDERT and colleagues (*The Lancet* **338**, 1471–1475; 1991) looked at data on 100 sets of twins born to women infected with HIV, and found that birth order and type of delivery had a considerable effect on which twin, if any, had become infected. In the cases that could be evaluated, 50 per cent and 38 per cent of first-born twins, delivered vaginally and by caesarian respectively, had been infected, against 19 per cent of second-born twins delivered by either route. The data support the idea that mother–infant transmission of HIV occurs not only during gestation but in labour; and the authors propose that they can be explained by the greater exposure of first-born twins (in caesarian deliveries, usually the one lying lower in the uterus) to infectious cervical and vaginal blood and mucus. They conclude that some form of cleansing the birth canal might well prove effective in reducing the incidence of mother–infant infection.

Two into nothing

PUT two oscillators with exactly the same natural frequency side by side, and you would expect to find perfect resonance. But not so. The resonant frequency of the pair is split into two components, one increased and the other decreased by a small amount that depends on the strength of coupling between the oscillators. F. Bernardot *et al.* (*Europhys. Lett.* **17**, 33–38; 1992) use as their resonators a collection of sodium atoms (as few as three) and an evacuated cylindrical cavity of superconducting niobium, each of which can absorb microwaves at a frequency of 68 gigahertz. As expected not only does the resonant frequency split into two (the Rabi splitting) as the atoms enter the cavity, but the splitting increases as more atoms are introduced. Similar experiments have been conducted at CalTech using just a pair of atoms in a Fabry–Perot resonator, and the test with a single atom is not far off.

Seeing red

IN 1986 Hogness and his colleagues identified three genes for the pigments of colour vision. Two of these were on the X chromosome and were assigned as the green and red cone pigments on the grounds that red–green colour blindness is X-linked. Now D. D. Oprian *et al.* (*Biochemistry* **30**, 11367–11372; 1991) have expressed and isolated all three proteins, which, when combined with their common chromophore, 11-*cis*-retinal, generate coloured products with absorption maxima at exactly the wavelengths found in single cones by microspectrophotometry. The spectroscopic mechanism by which the proteins induce such large and varied perturbations of the retinal spectrum is now open to study.

interpretation of the result would be that the postsynaptic response is increased.

So unless data arise that cannot be explained by present hypotheses on the mechanism of transmission in the central nervous system, an increase in the amplitude of miniature currents would, I believe, be interpreted most simply and by most workers in the field as a change in responsiveness of the postsynaptic membrane. Such a change has been seen by Manabe *et al.* As the authors observe, although the change in minia-

ture amplitude would be sufficient to explain the potentiation of evoked currents, additional presynaptic effects can certainly not be ruled out.

The Long Term Problem⁵ may not be completely solved. But this new study at least clarifies the point that an increase in postsynaptic response to release of a vesicle from the presynaptic terminal occurs on induction of LTP. □

Frances A. Edwards is in the Department of Pharmacology, University College London, Gower Street, London WC1E 6BT, UK.

MOLECULAR BIOLOGY**Extinction by indirect means**

Mary C. Weiss

FUSE two different types of somatic cell, and in the resulting hybrid the specialized gene products of one of the parent cells will often be extinguished. This phenomenon, known as tissue-specific extinction, is intriguing in itself, but study of it has also held out the promise of netting a very big fish indeed — an understanding of the direct negative-control mechanisms operating during the differentiation of mammalian cells.

Two groups, led respectively by G. Schütz (Boshart *et al.*¹) and R. E. K. Fournier (Jones *et al.*²), have now identified the molecular mechanism at play in one case of tissue-specific extinction. The first step on the way was made in 1984, when, using microcell fibroblast–hepatoma hybrids, Killary and Fournier³ described the chromosomal localization of the gene for TSE1 (tissue-specific extinguisher 1). They found that if chromosome 11 of the mouse (or its human homologue 17) from a fibroblast was introduced into a rat hepatoma cell genome, it caused extinction of expression of the liver-specific hormone-inducible enzyme tyrosine aminotransferase (TAT). Loss of this chromosome led to the re-expression of TAT. It was concluded that the chromosome carries a locus coding for a diffusible factor having a negative effect on expression of the TAT gene, from which it follows that the locus must be active in fibroblasts but not in hepatoma cells.

In subsequent work, Fournier and co-workers determined that TSE1 has several liver-specific target genes⁴, all of them inducible by glucocorticoids and cyclic AMP. A second important clue came from the observation that TSE1 action is fully alleviated by hormone action⁵. Boshart *et al.* found that the *cis*-acting sequences through which TSE1 action is mediated is a cAMP response element, located in the enhancer 3.6 kilobases upstream of the TAT gene. This deduction stemmed not only from

transfection assays involving the relevant sequences, but also from footprints *in vivo*⁶, and led Schütz's group to speculate that the TSE1 product changes the phosphorylation state of a protein important for functioning of the cAMP response element of the TAT gene (for example, by way of a specific phosphatase activity).

At this point it is helpful to recall the outline of the cAMP induction pathway. In the absence of hormonal stimulation, the enzyme protein kinase A is inactive because of the association of regulatory subunits with catalytic subunits. Cyclic AMP binds to the regulatory subunits to liberate the catalytic subunits. The catalytic subunits of protein kinase A can then phosphorylate a protein which binds to the cAMP response element, resulting in the activation of transcription. TSE1 could act directly or indirectly at any one of these steps.

The Schütz and Fournier laboratories achieved the molecular cloning of TSE1 in different ways. Both groups made use of microcell fibroblast–hepatoma hybrids containing only fragments of human chromosome 17 (ref. 7). These provide pairs of nearly isogenic clones where only a fragment of fibroblast chromosome 17q (marked by the *neo* gene) is retained, either including TSE1 or not. Boshart *et al.*¹ pursued their hypothesis that TSE1 prevents protein binding at the cAMP response element of the TAT gene. A first series of experiments revealed a sevenfold reduction in the activity of protein kinase A in TSE1⁺ compared to TSE1[−] hybrids, but only in the absence of cAMP. This implied that catalytic subunit expression is similar in the two cell lines. So Boshart *et al.* turned their attention to the regulatory subunits, thinking that perhaps one of them is up-regulated by TSE1. The answer was simpler than expected — Southern blot analysis revealed that regulatory subunit R1α is localized to

chromosome 17, and specifically to the fragment present in TSE1⁺ hybrids.

Meanwhile, Jones *et al.*² were using a subtracted-probe approach. Mapping of the chromosome 17 markers of the fragment containing hybrids permitted them to identify appropriate cell clones for this work: the TSE1⁺ cells chosen contained only 2–4 megabases of human DNA that were absent from a TSE1[−] clone. Labelled complementary DNA from the TSE1⁺ cells was subtracted with polyA⁺ RNA from the TSE1[−] cells, and the resulting single-stranded nonhybridizing material was used to screen a human fibroblast cDNA library. After various purification and verification steps, nine distinct human cDNAs were isolated. Mapping of the human cDNAs using a panel of ten fragment-containing microcell hybrid clones revealed that only one of them manifested concordant segregation with TSE1. Sequencing of the corresponding insert showed that it encoded the R1 α regulatory subunit of protein kinase A.

Both groups went a step further to demonstrate that TSE1 and R1 α are one and the same thing, transfection of R1 α cDNA into hepatoma cells producing a phenocopy of TSE1. In addition, Boshart *et al.*¹ showed that phosphorylation is reduced at serine-133 of the protein binding to the cAMP response element in TSE1⁺ cells, as well as in the R1 α transfectant.

How then does TSE1 (alias R1 α) produce its effect? Hepatoma cells produce very little R1 α : basal TAT activity would be the consequence of some constitutive activity of protein kinase A. Fibroblasts show high R1 α expression. The accrued synthesis of R1 α in TSE1⁺ hybrids would abolish basal activity, and consequently TAT expression, but this effect would be reversed upon addition of cAMP^{1,2}.

One issue that remains to be resolved, and is discussed in both papers, concerns the mechanism responsible for the maintenance of high (fibroblast) levels of R1 α expression when the chromosome is transferred to a hepatoma cell background, where endogenous R1 α expression is very low. The human fibroblast gene is not down-regulated to the hepatoma level, nor is the rat hepatoma gene up-regulated to the fibroblast level, im-

plying *cis*-heritability of the expression of this gene.

Jones *et al.*² conclude that "the TSE1 phenotype is not the consequence of repression; rather, it results from lack of activation". Some will be disappointed by this observation, in that extinction has not provided a handle to grasp direct negative-control mechanisms. But it does provide an answer to the worrisome question of why a fibroblast should produce 'repressors' for the entire set of

liver (or other differentiation-type) functions. The answer is that it does not. Only further work will reveal the mechanism underlying the effect of other tissue-specific extinguishers (the identification of which has proved more elusive than expected), but they may also turn out to act by indirect means. □

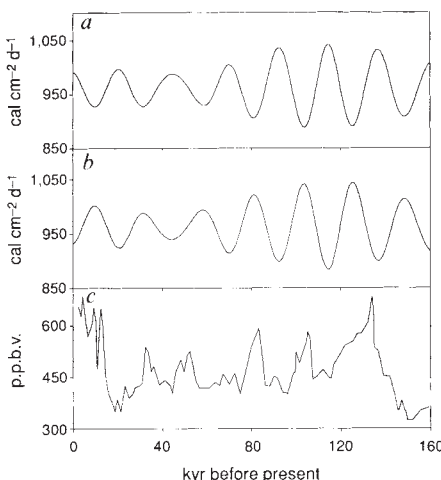
Mary C. Weiss is in the Department of Molecular Biology, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cédex 15, France.

ATMOSPHERIC METHANE

Tropical wetland sources

F. A. Street-Perrott

ONE of the changes in atmospheric composition currently causing greatest disquiet is the rapid rise in atmospheric methane (CH₄), a naturally occurring greenhouse gas which is 21 times more effective as an atmospheric-warming agent, molecule for molecule, than CO₂ (ref. 1). Data from the past 160,000 years, newly marshalled by Petit-Maire *et al.*², point to tropical wetlands as a leading influence on variations of atmospheric CH₄ levels.



Variations in solar insolation in summer, at 20° S (January, a) and 20° N (July, b), due to changes in the Earth's orbit over the past 160,000 years, and (c) in atmospheric CH₄ levels as recorded in Vostok ice⁵. (Insolation calculations by A. Berger, from US National Geophysical Data Center archive.)

Methane is emitted from a wide variety of anaerobic sources, including natural wetlands and freshwater bodies, rice paddies, enteric fermentation in ruminant animals, termite heaps, landfills and the oceans. It is also released inadvertently through biomass burning, coal mining, and oil and gas extraction. A small proportion, probably not even 1 per cent, may be being released owing to gradual warming of methane hydrates trapped in oceanic sediments and in the permafrost of Alaska, Canada and the Soviet Union³.

Methane is removed from the atmosphere primarily by reaction with hydroxyl radicals and by oxidation in dry soils³. Its concentration is now more than 1,700 parts per billion by volume (p.p.b.v.), and is increasing at a rate of 14–17 p.p.b.v. yr^{−1} (ref. 1). Popular concern has centred on the possibility that global warming might accelerate this rate of increase by destabilizing the enormous reserves of methane hydrates in the Arctic⁴.

Controversy

Chappellaz *et al.*⁵ set off a flurry of controversy by publishing data on CH₄ concentrations in bubbles of ancient air trapped in the Antarctic ice at Vostok Station. Over the past 160,000 years, CH₄ levels in the atmosphere have varied naturally from around 350 p.p.b.v. during glacial maxima to around 650 p.p.b.v. during warm interglacial periods, with sharp increases during the cold/warm transitions. It is clearly essential to understand these natural feedbacks before future trends can be predicted with confidence.

The CH₄ record from Vostok exhibits four significant periodicities — 110, 38, 24 and 19 thousand years (kyr) — in close agreement with the main components of the Earth's orbital variations: eccentricity (100 kyr), axial tilt (41 kyr) and the precession of the equinoxes (23 and 19 kyr). Chappellaz *et al.*⁵ argued that the global source must have increased by a factor of 2.3 during glacial/interglacial warming. In view of the strong precessional signal in the Vostok data, they proposed that orbitally driven changes in monsoon rainfall played a crucial role by controlling CH₄ emissions from low-latitude (tropical) wetlands. (Precession and eccentricity are the main controls over long-term variations in insolation in the tropics, whereas tilt becomes more important at high latitudes.)

Crowley⁶ took issue with this hypothesis, suggesting that a strong 23-kyr signal is also present in time series of

1. Boshart, M., Weih, F., Nichols, M. & Schütz, G. *Cell* **66**, 849–859 (1991).
2. Jones, K. W., Shapero, M. H., Chevrette, M. & Fournier, R. E. K. *Cell* **66**, 861–872 (1991).
3. Killary, A. M. & Fournier, R. E. K. *Cell* **38**, 523–534 (1984).
4. Lem, J., Chin, A. C., Thayer, M. J., Leach, R. J. & Fournier, R. E. K. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7302–7306 (1988).
5. Thayer, M. J. & Fournier, R. E. K. *Molec. cell. Biol.* **9**, 2837–2846 (1989).
6. Boshart, M., Weih, F., Schmidt, A., Fournier, R. E. K. & Schütz, G. *Cell* **61**, 905–916 (1990).
7. Leach, R. J., Thayer, M. J., Schafer, A. J. & Fournier, R. E. K. *Genomics* **5**, 167–176 (1989).

past temperatures in middle and high latitudes derived from both geological evidence and atmospheric modelling. Nevertheless, the new study by Petit-Maire *et al.*², who review the evidence for long-term variations in both wetland and dryland ecosystems in the Sahara and Sahel, provides further support for the tropical hypothesis. The most continuous and representative record of the extent of wetlands is provided by the concentration of freshwater diatoms (siliceous microfossils) in deep-sea sediments from the Equatorial Atlantic⁷. During arid phases, large quantities of diatoms were swept up from desiccated lakes and swamps and blown out to sea by the wind. In wetter periods, when the West African summer-monsoon rains were more intense and penetrated further north into the Sahara, the diatom-rich sediments were protected from erosion by water and vegetation.

The timing of the wet phases is supported by numerous geochronometric dates on outcrops of lake sediment in the Sahara and Sahel (although Petit-Maire *et al.* fail to point out that the interpretation of uranium-series dates on lacustrine materials is controversial). The four largest CH₄ peaks in the Vostok record also coincide with peaks in the discharge of the Nile, as recorded by sapropel (organic) layers in eastern Mediterranean deep-sea sediments.

Expanding wetlands

During periods of high rainfall, such as the early to mid-Holocene (10–5 kyr ago), freshwater habitats expanded dramatically across the region stretching from northern Algeria to Ghana in western Africa, and from southern Egypt to Tanzania in eastern Africa. In Mali alone, the area of swamps and shallow lakes increased by at least 56,000 km² (ref. 8). Much of the enlarged wetland area consisted of herbaceous (grass, sedge and reed) swamps⁹ with very high rates of CH₄ emission⁹. In central and eastern Africa, a key role was probably played by the tall sedge *Cyperus papyrus*, which grows in dense stands with very high primary productivity, rapid rates of decomposition, and highly reducing muds¹⁰. The northward spread of grassland, wooded grassland and forest during wet phases would also have greatly increased the biomass of wild herbivores and termites².

A further argument² for a strong tropical contribution to the Vostok record is the occurrence of a sharp decrease in atmospheric CH₄ (by about 170 p.p.b.v.) during the short cold episode known as the Younger Dryas at the end of the last glaciation. This corresponds to an important dry event in northern and eastern Africa¹¹, and northern South America¹². In contrast, there is at present little

evidence for a significant cold oscillation within the continental permafrost regions of North America and Eurasia¹³.

The arguments of Petit-Maire *et al.*, although persuasive, leave many questions unanswered. Recent estimates suggest that 24 (ref. 14) to 56 (ref. 9) per cent of the present CH₄ flux from wetlands originates between 30° N and 30° S. Natural wetlands still account for about a fifth of global CH₄ emissions³, although this figure must have been higher in the pre-agricultural era. Thus, it seems unlikely that orbitally driven changes in tropical precipitation alone can account for the scale of variations observed at Vostok, unless rates of emission from tropical sources are even higher than currently believed. More data on CH₄ fluxes from tropical ecosystems, especially dense papyrus stands, are urgently required. However, the palaeoclimatic record suggests that climate modellers must improve their predictions of tropical precipitation if the feedback effect of greenhouse warming on CH₄ emissions is to be reliably assessed.

A further difficulty arises because past insolation variations in the Northern and Southern Hemispheres were almost in antiphase (see figure). Although Petit-Maire *et al.* claim that the climatic evolution of the Sahara and Sahel was typical of the tropics at large, the interiors of southern Africa, Madagascar, and southern tropical and subtropical South America appear to have been relatively dry during the early and mid-Holocene^{15,16}. But it seems that these regions were much smaller and experienced weaker climatic changes than the vast areas influenced directly, or indirectly via the atmospheric circulation, by variations in Northern-Hemisphere insolation¹⁵. □

F. A. Street-Perrott is in the Environmental Change Unit and School of Geography, University of Oxford, Mansfield Road, Oxford OX1 3TB, UK.

1. Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. *Climate Change. The IPCC Scientific Assessment* (Cambridge University Press, 1990).
2. Petit-Maire, N., Fontugne, M. & Rouland, C. *Palaeogeogr. Palaeoclimat. Palaeoecol.* **86**, 197–204 (1991).
3. Cicerone, R. J. & Oremland, R. S. *Global biogeochem. Cycles* **2**, 299–327 (1988).
4. Nisbet, E. G. *Can. J. Earth Sci.* **26**, 1603–1611 (1989).
5. Chappellaz, J. *et al. Nature* **345**, 127–131 (1990).
6. Crowley, T. J. *Nature* **353**, 122–123 (1991).
7. Pokras, E. M. & Mix, A. C. *Quat. Res.* **24**, 137–149 (1985).
8. Street-Perrott, F. A. *et al. Trans. R. Soc. Edinb. Earth Sci.* **81**, 407–427 (1990).
9. Aselmann, I. & Crutzen, P. J. *J. Atmos. Chem.* **9**, 307–358 (1989).
10. Beadle, L. C. *The Inland Waters of Tropical Africa* 2nd ed. (Longman, London, 1981).
11. Street-Perrott, F. A. & Perrott, R. A. *Nature* **343**, 607–612 (1990).
12. Van der Hammen, Th. J. *Biogeogr.* **1**, 3–26 (1974).
13. Wright, H. E. *Quat. Sci. Rev.* **8**, 295–306 (1989).
14. Matthews, E. & Fung, I. *Global biogeochem. Cycles* **1**, 61–86 (1987).
15. COHMAP members *Science* **241**, 1043–1052 (1988).
16. Scott, L. *Palaeoecol. Afr.* **21**, 259–268 (1990).

Second sight

IN three-dimensional television, two cameras view the same scene. Each relays its image to one eye of the viewer, who fuses the images into a subjective picture with stereoscopic depth. Daedalus has devised a stereo-TV helmet to do this in real time. Each of the eyepieces is fed from its own forward-pointing single-chip TV camera; the wearer sees the world around him as a binocular TV image. The cunning part is this: an electronic frame-store slightly delays the image received by one eye. Daedalus wants to find out how much delay is needed to prevent the stereoscopic fusion of a moving scene, and whether viewers can learn to tolerate and interpret scenes with a significant time-lag between left and right images.

The applications are obvious. Our senses judge rates of change very poorly. Many motor accidents, for example, are caused by bad guesses of the closing rate of another vehicle. But a driver wearing Daedalus's TV time-lag helmet will experience rates of change directly, as a difference in size or position between left and right images of an object. He will have wonderfully acute reactions, and should make a superlative driver, test-pilot or astronaut.

More cunning still, Daedalus points out that the left-hand part of the visual field goes directly to the right lobe of the brain, which is intuitive and nonverbal. The right-hand part of the scene goes to the left lobe, which handles speech and reasoning. So Daedalus is modifying his TV time-lag helmet to delay the right-hand half of the picture presented to each eye with respect to the left-hand half (or vice versa).

The psychological impact should be profound. Although the left and right brain lobes generally pool their knowledge, even a slight time-difference between them would give a powerful advantage to the lobe that gets the information first. A normally logical wearer could develop his intuitive side, while a wearer dominated by immediate emotional reactions could delay them, and see things more coolly and rationally. One side effect might be a strong sense of *déjà vu*, that weird feeling of having been in exactly this situation before. This is claimed to arise when the two brain lobes get out of synchronism, so that one is seeing what the other has already stored as memory. Some psychiatric patients feel this way almost all the time. A suitable reverse setting of Daedalus's delay helmet should be able to cancel the effect, and restore the sufferer to real-time existence. *Déjà vu* would at last be definitively explained, and the inter-lobal time-lag that causes it could be measured.

David Jones

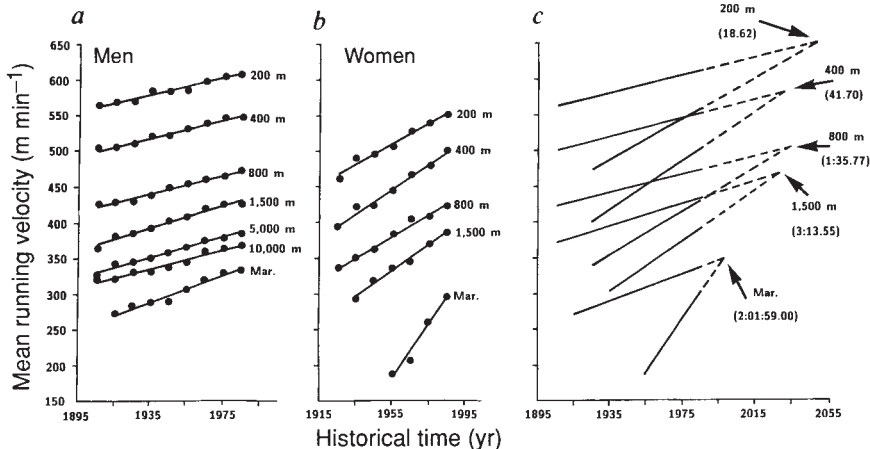
Will women soon outrun men?

SIR — The mean running velocity ($\bar{V}$) is a crucial determinant of the metabolic demands imposed by competitive running events^{1,2}. Its historical progression, therefore, is likely to be important in understanding the physiological determinants of the seemingly inexorable progression of record performances.

We therefore established $\bar{V}$ as a function of historical time (t) for the world records at all the standard Olympic events from the 200 m to the marathon

no different for men and women within the first half of the twenty-first century (c in the figure). Beyond that time, current progression rates imply superior performance by women. The projected intersection for the marathon is 1998.

The suggestion that women could, so soon, be running these races as fast as men seems improbable at first appearance. None of the current women's world record holders at these events could even meet the men's qualifying



World record progression, expressed as mean running velocity versus historical time, for men (a) and women (b), with best-fit linear regressions (solid lines) superimposed. In c, the regression lines for the common events for men and women (solid lines) are extrapolated (dashed lines) to their points of intersection; the predicted world record times at these intersection points are shown in parentheses (h:min:s)

(42,195 m) for men, decade-by-decade, throughout this century^{3,4}. We were able to establish this relationship only for events up to 1,500 m since the 1920s for women; data were inadequate for the 5,000 and 10,000 m, although we judged there to be sufficient for the marathon.

In men, the progression of $\bar{V}$ appears to be a linear function of t , with slopes for the different events being remarkably similar (a in the figure) — in agreement with the results of Ryder *et al.*⁵. These ranged from 5.69 to 7.57 m min⁻¹ decade, with no systematic variation with increasing race distance. The marathon slope, however, was appreciably greater (9.18 m min⁻¹ decade).

For women, there were also no significant differences in the slopes among the different events up to the 1,500 m (b in the figure). The slope, however, was approximately double that for the men, ranging from 14.04 to 17.86 m min⁻¹ decade. As for men, the rate at which $\bar{V}$ increased in the marathon was appreciably greater (37.75 m min⁻¹ decade). Despite the potential pitfalls, we could not resist extrapolating these record progressions into the future.

Unless the progression rate of men's records increases relative to that of women, then $\bar{V}$ for these events will be

standard to compete in the 1992 Olympic games. However, it is the rates of improvement that are so strikingly different — the gap is progressively closing⁶.

Although it is difficult to establish a precise metabolic energy equivalent of these rates of improvement, one may estimate from the data of Margaria *et al.*⁷ and Wyndham *et al.*⁸ that, for the events up to 10,000 m, the progression requires a rate of increase in O₂ consumption of about 10 ml min⁻¹ yr for men and more than double that for women.

It is unlikely that we will learn when, and how rapidly, the current high rate of improvement was established, owing to the lack of reliable times over reliable distances in the past. Some world records are available however, as far back as the 1860s (ref. 4); they are consistent with the values 'expected' from the current progression rates. Whether the world record progression rate will begin to slow, either relatively abruptly or more progressively, will only become apparent in the future.

In any event these results pose four challenging questions to physiologists. Why is: the world record progression in the various events so linear over an interval of approximately a century; the

slope of the record progression so similar from the sprints to the 10,000 m; the record progression in the marathon appreciably greater; and the record progressions for women increasing at such a rapid rate relative to men?

BRIAN J. WHIPP
SUSAN A. WARD

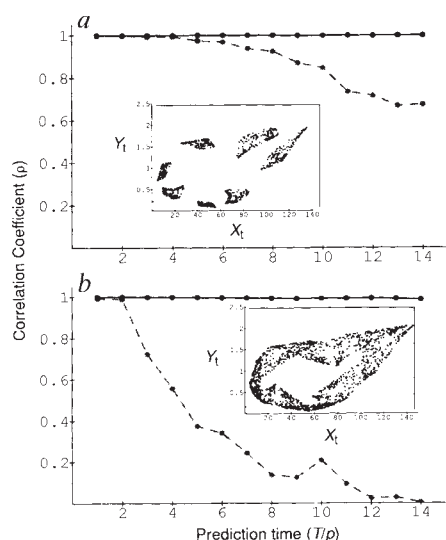
Departments of Physiology and
Anesthesiology,
School of Medicine,
University of California,
Los Angeles, California 90024, USA

1. Lloyd, B. B. *Circ. Res.* **XX** & **XXI** (suppl. 1), 1218–226 (1967).
2. Di Prampero, P. E. *Rivista di Cultura Sportiva* **3**, 3–7 (1984).
3. Matthews, P. *Track and Field Athletics: The Records* (Guinness, Enfield, 1986).
4. *Progression of World's Best Performances and Official I.A.A.F. World Records* (International Athletic Foundation, Monaco, 1987).
5. Ryder, H. W., Carr, H. J. & Herget, P. *Sci. Am.* **234**, 6 109–119 (1976).
6. Dyer, K. J. *J. Biosocial Sci.* **9**, 325–338 (1977).
7. Margaria, R., Cerretelli, P., Aghemo, P. & Sassi, G. *J. appl. Physiol.* **18**, 367–371, (1963).
8. Wyndham, C. H., Strydom, N. B., Van Rensburg, A. J. & Benade, A. J. *S. Afr. Med. J.* **43**, 996–1002 (1969).

How predictable is chaos?

SIR — Sugihara and May¹ have developed a novel approach to distinguish chaos from noise in short time series. The principle lies in a geometric method designed to make predictions based on a library of past patterns in the series, with a view of comparing them with actual trajectories. The necessity for such a technique arises in fields like population biology, where data are typically sparse. Without questioning the basis of the approach, we point out that limitations exist: the method fails to detect chaos in series of data simulated by a simple and general age-structured population model.

One reason is that corresponding strange attractors are split into several disconnected pieces. The iterations of any point chosen in one of the fragments are located successively in the other pieces (in a precise order), before falling back into the initial one, and so on. As a consequence, there is an overall period in the structure of the attractors, which is equal to the number of pieces. It has been known for some time that this pattern occurs in discrete systems when the transition towards chaos runs through an invariant circle that is kinked². Such bifurcations are common in ecological models (see refs 3 and 4). They compromise attempts to identify chaotic dynamics with the method of Sugihara and May, which is influenced by the overall periodicity of the attractors, whereas chaos is hidden in what may well be classified as 'residual noise'. More surprisingly, this difficulty still per-



Analysis of a chaotic time series simulated by the density-dependent two age-classes model: $x_t = r \cdot y_t \exp(-0.01 \cdot (y_t + x_t))$, $y_t = 0.2 \cdot x_t \exp(-0.01 \cdot (y_t + x_t)) + 0.8 \cdot y_t \exp(-0.05 \cdot (y_t + 0.5 \cdot x_t))$. *a*, $r = 112.5$; *b*, $r = 117$. Small boxes, the corresponding attractors; solid lines, analysis of the full sequence $(x_t)_{0 \leq t \leq 1,000}$; dashed lines, analysis of the sub-sequence $(x_{7,i})_{0 \leq i \leq 1,000}$.

sists when attractors are no longer fragmented (this is obtained by tuning one parameter of the model). Now the overall periodicity is far less obvious, but probably responsible again for our unexpected results.

These phenomena are illustrated in the figure. In *a*, the overall period of the strange attractor is 7, and chaos is not detected on 1,000 consecutive points: the correlation coefficient between predictions and actual data points shows no dependence on the prediction-time interval. However, restricting the analysis to a sub-sequence of points located in one of the seven pieces of the attractor provides the signature of chaotic dynamics¹. In *b*, a similar comparison is performed; the conclusions are identical, although the overall periodicity is not evident.

Supported by recent analyses of models for single-species populations (R. H. F. and M. Gatto, in preparation) or multiple-species food chains⁵, there is an increasing belief that chaos may commonly exist, in natural systems, in a form which appears to be regular oscillations over short timescales⁴. Therefore, to improve the applicability of the Sugihara and May method, one should first

test any time series $(X_t)_{t \geq 0}$ for an eventual cyclic trend (with period τ), and then look for chaos in the sub-sequence $(X_{t,i})_{i \geq 0}$. (However, this precaution generates additional need for a rather long time series, all the more so as τ increases.) Real data previously interpreted as reflecting noisy cycles should be reconsidered in this way.

BERNARD CAZELLES

Laboratoire de Biomathématiques,
INSERM U 263,
Université Paris 7, 2 place Jussieu,
75251 Paris Cedex 05, France

REGIS H. FERRIERE*

Laboratoire d'Ecologie, CNRS URA 258,
Ecole Normale Supérieure,
46 rue d'Ulm,
75230 Paris Cedex 05, France

*To whom correspondence should be addressed.

SUGIHARA AND MAY REPLY — One problem in trying to reconstruct the attractor that may underlie an apparently chaotic time-series is that we often have data for only one variable in what is, intrinsically, a multivariate system. Following Takens (F. Takens, *Lect. Notes Math.* **898**, 366–381; 1981), we may 'embed' our one-dimensional time series in an E -dimensional phase space by, in effect, running a fork with E tines, spaced at intervals τ , along the time series. By regarding the E data points the tines touch as a single E -dimensional point, we generate a constellation of points in E -space.

The difficulty is that, in general, we have no *a priori* knowledge about the proper values of E and τ ; that is, about the number of tines and their spacing. Very often, commonsense has resulted in the data being collected at intervals that roughly correspond to the appropriate value of τ , so that the difficulties primarily lie in finding the best E . Mainly for this reason, our paper focused on the embedding dimension E (the number of tines on the fork); although we did refer to the question of appropriate τ -values (spacing between the tines; see our Fig. 3*b*), we gave it little emphasis. Cazelles and Ferrière make the useful point that appropriate spacing — appropriate τ -values — can be important.

GEORGE SUGIHARA

ROBERT M. MAY

Department of Zoology,
University of Oxford,
Oxford OX1 3PS, UK

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references. □

Changes in Earth rotation rate

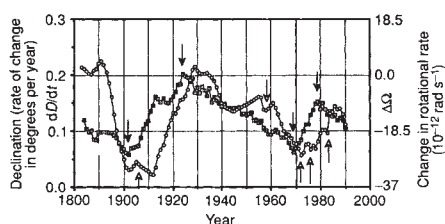
SIR — Variations of the Earth's rotation rate with periods less than two years are due to variations in the zonal circulation of the upper atmosphere. The longer-period variations, called decade variations, are attributed to exchanges of angular momentum between the Earth's rigid mantle and its liquid core. Ten years ago, we proposed¹ the existence of a significant correlation between these decade fluctuations of the Earth's rotation rate and those of a properly chosen indicator of secular variation of the geomagnetic field. Based on the observation of a secular variation impulse in 1969–70, we predicted² that an acceleration of the rotation rate would take place in the early 1980s. This indeed was what happened³. Observation at that time of more recent secular variation (a new sharp change of slope in 1978) led to a second prediction of a deceleration in rotation rate starting about ten years later³. We can now confirm this prediction and the 9 ± 2 -year lag between the two data sets (see figure).

In 1981, we suggested that the correlation was due to electromagnetic coupling between the core and mantle, the source of this coupling lying with the westward bodily drift or rotation of the external layers of the core with respect to the mantle. Much work has been done since then to study the flow at the core–mantle boundary. Westward drift appears now to be limited to the Atlantic hemisphere and has much less significance than previously thought; it is now even doubtful whether any body rotation exists at all⁴.

Electromagnetic coupling is, on the other hand, faced with very significant difficulties in accounting for the observed very rapid slope changes in the Ω curve⁵. The favoured physical mechanism appears to be topographic core–mantle coupling, first proposed by Hide⁶. If the core–mantle boundary displays topography (departures from axial symmetry), flow at the surface of the core exerts a net torque on the mantle, except for very special flow configurations⁷.

If the amplitude of the topography reaches a few hundred metres, as inferred from nutation data, or even larger, as proposed by some seismologists, the calculated torque is found to be orders of magnitude larger than allowed by the rotation data (for example, ref. 8). One must then assume that flow in the upper core remains on average locked in a specific configuration with respect to topography of the core–mantle boundary such that the resulting net torque is far

1. Sugihara, G. & May, R. M. *Nature* **344**, 734–741 (1990).
2. Curry, J. H. & Yorke, J. A. in *The Structure of Attractors in Dynamical Systems. Lect. Notes in Math.* **668**, 48–66 (Springer, New York, 1978).
3. Schaffer, W. M. & Kot, M. *Trends Ecol. Evol.* **1**, 58–63 (1986).
4. Ellner, S. in *Mathematical Approaches to Problems in Resource Management and Epidemiology* (eds Castillo-Chavez, C., Levin, S. A. & Schoemaker, C. S.) *Lect. Notes in Biomath.* **81**, 286–307 (Springer, New York, 1989).
5. Hastings, A. & Powell, T. *Ecology* **72**, 896–903 (1991).



Changes in the Earth's rotation rate (circles, $\Delta\Omega$) in 10^{-12} rad s^{-1} (after ref. 9) are compared with secular variation of declination (squares, rate of change dD/dt) in degrees per year at the French national observatory (1883–1901, St Maur; 1902–1935, Val Joyeux; 1936–1990, Chambon-la-Forêt), considered as a representative indicator of geomagnetic secular variation. Open arrows, times of large Southern atmospheric oscillations; closed arrows, geomagnetic secular variation sharp changes of slope (around 1900, 1923, 1969 and 1978), otherwise called impulses or jerks. The fact that the time derivative of a single field component at a single observatory does provide a good indicator of worldwide secular variation (and in particular of jerks) has been discussed elsewhere, and illustrated by a number of harmonic coefficients of global expansions (see ref. 10). The normalized cross-covariance between the two curves reaches a clear, principal maximum of 0.74 for a time offset of 9 ± 2 years.

smaller than predicted by simple order-of-magnitude considerations. Small variations of the flow with respect to this configuration can, however, take place; if at some time a change in flow configuration suddenly occurs, this will be translated instantaneously into a change in the rate of secular variation (except for the delay introduced by the conducting mantle, which is probably of the order of a year). The torque exerted on the mantle will increase (from zero) and, when large enough will, with some delay, accelerate or decelerate the mantle.

J. L. LE MOUËL
V. COURTILLOT
D. JAULT

*Institut de Physique du Globe,
Département de Géomagnétisme et
Paléomagnétisme,
4 Place Jussieu,
75252 Paris Cedex 05, France*

1. Le Mouél, J. L., Madden, T. R., Ducruix, J. & Courtillot, V. *Nature*, **290**, 763–765 (1981).
2. Courtillot, V., Ducruix, J. & Le Mouél, J. L. *C.R. Acad. Sci. Paris. Ser. D*, **287**, 1095–1098 (1978).
3. Gire, C. & Mouél, J. L. in *The Earth's Rotation: Solved and Unsolved Problems* (ed. Cazenave, A.) 241–258 (Reidel, Dordrecht), (1986).
4. Jault, D., Gire, C. & Le Mouél, J. L. *Nature*, **333**, 353–356 (1988).
5. Jault, D. & Le Mouél, J. L. *J. geomagn. Geoelectr.*, **43**, 111–129 (1991).
6. Hide, R. *Nature*, **222**, 1055–1056 (1969).
7. Jault, D. & Le Mouél, J. L. *Geophys. J. Int.*, **101**, 233–241 (1990).
8. Hinderer, J., Legros, H., Jault, D. & Le Mouél, J. L. *Phys. Earth planet. Int.*, **59**, 329–341 (1990).
9. Morrison, L. V. *Geophys. J. R. astr. Soc.*, **58**, 349–360 (1978).
10. Langel, R. A. & Estes, R. H. *J. geophys. Res.*, **90**, 2495–2509 (1985).

Anion guests in heteropolyanions?

SIR — Müller¹ and Day *et al.*², stimulated by Mitchell's News and Views article on inorganic host-guest chemistry³, cited heteropolyanions as examples in which both anions and cations could be guests (or 'hostages'). I can claim some familiarity⁴ with this remarkably versatile class of compounds, which is also of growing significance in other disciplines such as materials science, medicine and topology.

In my view the 'anion-centred' heteropolyanions like $[PMo_{12}O_{40}]^{3-}$ are not analogous to the novel polyvanadate clathrate species that are being explored by Müller *et al.* and which were discussed by Mitchell. The host cluster shells of the latter species are negatively charged (and encapsulate anions) and are constructed of $OV^{IV}O_4$ or $OV^{V}O_4$ square pyramids. Simple valence-sum considerations imply that there will be very little excess charge density available for conventional V...O bonding in a direction trans to the terminal oxo groups, that is, towards the interior of the shell. Although cluster shells of $OMo^{VI}O_4$ or $OW^{VI}O_4$ square pyramids are also implicit in the clathrate description of ions like $[PMo_{12}O_{40}]^{3-}$ and

$[PW_{12}O_{40}]^{3-}$, such shells are uncharged, and the higher valences of molybdenum or tungsten *vis-à-vis* vanadium result in stronger interactions with the oxygens of the internal oxoanion. This leads to the conventional descriptions of these structures in terms of OMO_5 octahedra.

Furthermore, it is clear from the observed structural variety of polyvanadate shells that the encapsulated anions (which are not even restricted to oxoanions for these complexes) act as templates and control the organization, size and shape of the surrounding cluster shells. This feature stands in contrast to the archetypal heteropolyanion Keggin structure which can be generated in the absence of tetrahedral oxanion templates, for example, the metatungstate anion $[H_2W_{12}O_{40}]^{6-}$.

MICHAEL T. POPE

*Department of Chemistry,
Georgetown University,
Washington, DC 20057, USA*

1. Müller, A. *Nature*, **352**, 115 (1991).
2. Day, V. W., Klemperer, W. G. & Yaghi, O. M. *Nature*, **352**, 115–116 (1991).
3. Mitchell, P. C. H. *Nature*, **348**, 15–16 (1990).
4. Pope, M. T. *Heteropoly and Isopoly Oxometalates* (Springer, Berlin, 1983).

No lubricants from fluorinated C_{60}

SIR — Ever since C_{60} was discovered¹, there has been anticipation of the possible use of the fluorinated derivatives, and $C_{60}F_{60}$ in particular, as lubricants. There are indications, however, that, unlike Teflon, this material might not be chemically stable: fluorination of C_{60} is very slow, and steric crowding is predicted to increase the C–C bond length by about 0.05 Å (ref. 2) and decrease the C–F bond strength by 15% relative to that in CF_4 (ref. 3). We have found that highly fluorinated C_{60} does indeed seem to be highly chemically reactive.

In a study of the fluorination of C_{60} to give a mixture of fluorinated derivatives⁴, we ascribed a peak at -188.5 p.p.m. in the ^{19}F NMR spectrum of the product to HF, the origin of which was unclear. Moreover, the acetone solvent underwent rapid aldol condensation and darkening, consistent with the presence of HF. Handling the fluorinated material in air, however, seemed to present no great difficulties.

In further studies, we exposed C_{60} (contained in a Teflon tube rather than in glass as used previously) to fluorine gas for about six weeks. Acetone solutions of samples of the product discoloured immediately and the ^{19}F NMR spectra showed a very substantial peak at -190.5 ± 1.5 p.p.m.; this peak also appeared in a tetrahydrofuran (THF)

solution, and its position and intensity varied with the amount of water present in the solvent. In confirming that this peak is due to HF, we find that fluorinated C_{60} undergoes very rapid reaction with sodium carbonate to give sodium fluoride and a yellow water-soluble product, the colour intensity of which differs for acid and basic solutions (being deeper in the latter). We have monitored this reaction via disappearance of the broad band⁴ in the ^{19}F NMR spectrum. This band decreased in intensity (at a decreasing rate) over 48 h, and could then be further decreased by addition of sodium hydroxide. Fluorinated C_{60} undergoes nucleophilic substitution by hydroxide (confirmed by infrared studies), and the rate of this reaction decreases as the fluorine content of the fullerene becomes reduced.

The apparent stability of the material in air is presumably a consequence of its hydrophobic nature, which renders it relatively stable and insoluble. A sample suspended in water initially showed no visible signs of reaction, but on addition of a few drops of THF, an immediate and exothermic reaction took place, and NMR analysis showed that substantial quantities of HF were produced. The formation of HF in acetone or THF solutions is thus due to the presence of traces of water, which attacks the solu-

bilized compound rapidly.

These findings rule out the use of highly fluorinated C_{60} as a lubricant because, while attack of the solid form by water is slow, it will occur over time to release HF. They also account for the fact that the fluorine content of crystalline fluorinated C_{60} , tentatively attributed by us⁴ to $C_{60}F_{60}$, decreased gradually on standing in air, with accompanying etching of the container: slow nucleophilic substitution by atmospheric moisture occurs to release HF. Other halogenated C_{60} derivatives may be even more reactive, in which case care may be necessary in handling them. We are now

investigating the synthetic use of reactions of fluorinated C_{60} with a wide range of nucleophiles.

ROGER TAYLOR

ANTHONY G. AVENT

T. JOHN DENNIS

JONATHAN P. HARE

HAROLD W. KROTO

DAVID R. M. WALTON

School of Chemistry and Molecular Sciences,

University of Sussex,

Falmer, Brighton BN1 9QJ, UK

JOHN H. HOLLOWAY

ERIC G. HOPE

Department of Chemistry,

University of Leicester,

University Road,

Leicester LE1 7RH, UK

G. JOHN LANGLEY

Department of Chemistry,

University of Southampton,

Southampton SO9 5NH, UK

1. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–164 (1985).

2. Scuseria, G. E. *Chem. Phys. Lett.* **176**, 423 (1991).

3. Dunlap, B. I., Brenner, D. W., Mintmore, J. W., Mowrey, R. C. & White, C. T. *J. phys. Chem.* **95**, 5763–5768 (1991).

4. Holloway, J. H. et al. *J. chem. Soc. Chem. Commun.* 966–969 (1991).

Strontium isotopes at K/T boundary

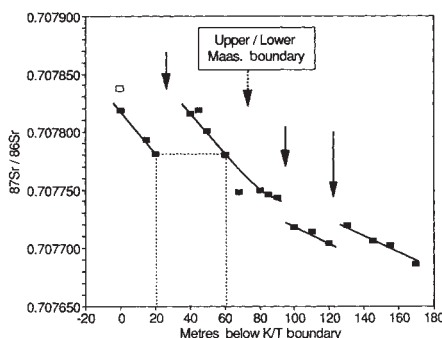
SIR — Nelson *et al.*¹ attempt to bring needed objectivity to an examination of the evolution of marine $^{87}\text{Sr}/^{86}\text{Sr}$ near the Cretaceous/Tertiary (K/T) boundary. In our view, the complexity of their section, together with conflicts within their own biostratigraphy, invalidate their treatment. The calculation of $\Delta^{87}\text{Sr}/^{86}\text{Sr}$ Myr⁻¹ for the section at Bidart is possible only if numerically dated levels can be accurately placed in the section and the $\Delta^{87}\text{Sr}/^{86}\text{Sr}$ between them accurately assessed. Neither has been accomplished for the following reasons.

Nelson *et al.* calculated $\Delta^{87}\text{Sr}/^{86}\text{Sr}$ Myr⁻¹ for part of Maastrichtian time. Critical to this calculation is the assignment of a date of 2.5 Myr before the K/T boundary to the first appearance of *Abathomphalus mayaroensis*. This date derives from a zonal diagram in ref. 2. Derived from ref. 3, this date is 1 Myr. Neither figure is reliable, as no radiometric markers for these horizons are known to us. Furthermore, their placement of the base of the *A. mayaroensis* zone conflicts with their own nannofossil zonation. The zone of the planktonic foraminiferid *A. mayaroensis* is commonly equated with the upper part of the Upper Maastrichtian. The calcareous nannofossil zone of *L. praequadratus* is usually placed in the Lower Maastrichtian. On Fig. 1 of ref. 1, bases of the zones almost coincide (108 and 110 m). The nannofossil evidence is more reliable, and it identifies the section around 100 m as Lower Maastrichtian, not Upper Maastrichtian, as Nelson *et al.* claim. The Upper Maastrichtian probably starts around 70 m at the base of the *L. quadratus* zone, and it is at about 70 m that Late Maastrichtian

macrofossils first appear⁴.

Nelson *et al.* assume that their composite section is complete. But Clauser⁵ remarks “*Toutefois les biozones peuvent être légèrement incomplètes, du fait de contacts tectoniques mineurs*”. Furthermore, the parts of the composite section of Nelson *et al.* are not “separated by a fault” but by numerous faults within 0.5 km of intervening, folded, section⁵. In addition, all parts are extensively faulted.

In the figure we offer an alternative interpretation of the data of Nelson *et al.* The three Cretaceous samples from the upper part of the composite section define a $\Delta^{87}\text{Sr}/^{86}\text{Sr}$ per metre of section that is equal to that shown by samples between 40 and 60 m from the lower part. It seems likely that strata in these parts of the section have been mismatched and repetition has occurred, with the ‘Danian’ sample being merely the highest point in a tectonically interrupted



Alternative interpretation (heavy lines) of the data of Nelson *et al.*¹, plotted as in their Fig. 1. Open square, ‘Danian’ sample; dotted lines, extent of overlap of the two parts of the composite section; solid arrows, positions of apparent geological breaks in the section (refs 1 and 5).

upward trend of $^{87}\text{Sr}/^{86}\text{Sr}$. In the figure the gradients of the two parts of the section are used to measure the overlap, which is 40 m. The figure also shows the position of apparent breaks in the section⁵, which in our interpretation seem to coincide with isotope discontinuities. Furthermore, there was a marked decrease in sedimentation rate in the area between Early and Late Maastrichtian times^{4,5}. We suggest this caused the break-in-slope at about 70 m shown by our figure. An Early Maastrichtian age for the section immediately below 70 m is supported by the occurrence of *Pachydiscus neubergicus*⁴, which is predominantly Early Maastrichtian.

A calculation of $\Delta^{87}\text{Sr}/^{86}\text{Sr}$ Myr⁻¹ requires acceptance of some assumptions, foremost of which is that the Campanian/Maastrichtian (C/M) boundary can be located in the section. Macropalaeontological criteria will place it at a higher level than will micropalaeontological criteria. Our interpretation of the non-standard nannofossil zonation of Nelson *et al.* places the nannofossil C/M boundary at about 130 ± 30 m below the K/T boundary (our unpublished nannofossil zonation places it at about 160 m). The macrofossil C/M boundary must be below the first appearance of *P. neubergicus* at about 125 m². We adopt a figure of 130–140 m for the approximate level of the macrofossil C/M boundary. Further assumptions are that the data of Nelson *et al.* show that $\Delta^{87}\text{Sr}/^{86}\text{Sr}$ between the K/T and C/M boundaries is about 110×10^{-6} , and that the numerical ages of these boundaries are $66 \pm$ and 72 ± 2 Myr, respectively. If one accepts these assumptions, the mean $\Delta^{87}\text{Sr}/^{86}\text{Sr}$ Myr⁻¹ during Maastrichtian time is between 11 and 55×10^{-6} Myr⁻¹, with a mean of 18×10^{-6} Myr⁻¹, which is 15% of the rate claimed by Nelson *et al.* and less than the long-term increase of $25 \pm 5 \times 10^{-6}$ Myr⁻¹ that occurred during most of the Late Cretaceous⁶.

J. M. MCARTHUR

J. BURNETT

Department of Geological Sciences,
University College London,
Gower Street,
London WC2 6BT, UK

J. M. HANCOCK

Department of Geology,
Royal School of Mines,
Imperial College,
London SW7 2BP, UK

1. Nelson, B. K., MacLeod, G. K. & Ward, P. D. *Nature* **351**, 644–647 (1991).

2. Kent, D. & Gradstein, F. G. *Bull. Geol. Soc.* **96**, 1419–1427 (1985).

3. Haq, B. U., Hardenbol, J. & Vail, P. R. *Science* **235**, 1123–1167 (1987).

4. Ward, P. D. *Revista Espanola de Paleontol. Extraord.* **119**, 126 (1988).

5. Clauser, S. C. R. *Acad. Sc. Paris* **304**, ser. II, 11, 579–584 (1987).

6. McArthur, J. M. et al. *Geol. Soc. Spec. Pub. High Resolution Stratigraphy* (eds Hailwood, E. & Kid, R.) (in the press).

Solar conundrums

Richard D. Petrasso

Blinded by the Light: The Secret Life of the Sun. By John Gribbin. Bantam: 1991. Pp. 226. £14.99.

THE Sun, from ancient to modern times, has been pre-eminent in developing man's understanding of the Universe. In the fifth century BC, Anaxagoras showed it to be a 35-mile-diameter ball of fiery stone orbiting 4,000 miles overhead. In the late nineteenth century, Lord Kelvin (William Thomson) limited its age to a mere 25 million years. And today, exotic particles called WIMPs are hypothesized to orbit within its core. Of course, Anaxagoras's and Lord Kelvin's conclusions proved to be far off the mark, yet both men made mathematically impeccable calculations based on the best though flawed assumptions of their eras. And what about today's WIMPs? Are they, as claimed, the basis of a solution to the outstanding solar neutrino problem?

Such issues and questions form the grist of *Blinded by the Light*. By focusing on the conundrums of the Sun's energy production and age, Gribbin paints an engaging picture of how science evolves, pivotal false starts and all. And for those with an interest in the dramatic interplay between physics, biology and geology in the nineteenth century, no more-illuminating episode exists than the controversy over the age of the Sun and the Earth.

This controversy, as Gribbin notes, pitted Charles Darwin against Lord Kelvin. Darwin was greatly influenced by Charles Lyell's 1830 treatise on the immeasurably vast times needed for geological evolution and by the immense timescale needed for his own theory of biological evolution. On this basis, he qualitatively estimated that the denudation of the chalk cliffs of the Weald in Kent took roughly 300 million years and, therefore, that the age of the Earth was much greater. By contrast, Kelvin, having quantified Herman von Helmholtz's hypothesis of heat generation by way of nebular gravitational condensation, obtained 20 million years as the probable duration of the Sun's energy production. In an article published in *Macmillan's Magazine* in 1862, Kelvin belittled Darwin's estimate:

What then are we to think of such geological estimates as 300 million years for the 'denudation of the Weald'? Whether it is more probable that physical conditions of the sun's matter differ 1000 times more than dynamics compel us to suppose they differ from those of matter in our laboratories; or that a stormy sea, with possibly channel tides of extreme violence, should encroach on a chalk cliff 1000 times more rapidly than Mr. Darwin's estimate of one inch per century?

Through this and other quotations and discussion, Gribbin effectively highlights this debate. But I disagree with his view that Darwin's wealden estimate was "rather careless". This misses a fundamental point. In a subsection of the first edition (1859) of *The Origin* entitled "On the Lapse of Time...", Darwin admirably and severely reined in Lyell's immeasurably vast times by intelligently discussing the accumulation and denuda-

tion of geological features, ending with the wealden estimate that he freely admits is both "crude" and "highly imperfect". Nonetheless, this was an original accomplishment and a serious effort to establish a concrete timescale based on a geological record. (20 million years is the upper limit for a present-day estimate of the wealden denudation.)

So why did Kelvin so high-handedly attack Darwin's estimate? Part of the answer is rooted in the seductive rigour afforded by his own calculations, and the lack thereof in Darwin's. But undercurrents in Kelvin's written work indicate still more was involved. Specifically, Kelvin goes on to admit in his 1862 article that, owing to material uncertainties, an extreme upper limit of 500 million years exists for the solar duration. From this and other inconsistencies, I am left with the impression that Kelvin was annoyed by Darwin's *Origin* because he thought it challenged his own deep-rooted belief in the ordained status of the "race of intelligent beings".

It is unfortunate that neither Gribbin

nor most other recent writers on the history of solar science consider the seminal contributions of the physicist Homer Lane. While Kelvin and Helmholtz were treating the Sun as a white-hot fluid of generally unknown properties unamenable to detailed analysis, Lane published in 1870 in the *American Journal of Science* a detailed calculation of the Sun's temperature and density. By assuming the validity of the perfect gas law and hydrostatic equilibrium, Lane obtained a central temperature and density of the order of 20×10^6 K and 30 g per cm³, a stunning achievement considering today's values of 15×10^6 K and 150 g per cm³. This is all the more

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The conundrum over the age of the Earth and Sun pitted Kelvin (left) against Darwin. Because Darwin's theory of evolution required vast periods of time, Darwin was increasingly pained by the apparent invincibility of Kelvin's shrinking estimates for the solar duration, which eventually converged to about 25 million years.

impressive when one considers that Lane could not have known that the Sun consists of a plasma of electrons and ions also subject to the perfect gas law and hydrostatic equilibrium. Even Kelvin was struck by the "great power" of Lane's calculations. Nevertheless, in his 1887 address ("On the Sun's Heat") to the Royal Institution of Great Britain, Kelvin argued that Lane's estimate of the central density (but not temperature) should be "much reduced". (Here Gribbin also errs in claiming that in that address Kelvin did not properly credit Helmholtz for the condensation model.)

Gribbin then goes on to describe how Henri Becquerel's revolutionary discovery in 1896 of radioactivity demolished Kelvin's age estimates for both the Sun and the Earth. Savour the irony of subsequent developments when the astrophysicist George Darwin, son of Charles and nonetheless younger col-

Mary Evans

league of Kelvin, conjectured in 1903 in this very journal:

We have recently learnt the existence of another source of energy, and that the amount of energy available is so great as to render it impossible to say how long the sun's heat has already existed, and how long it will last in the future . . . I think we have no right to assume that the sun is incapable of liberating atomic energy to a degree at least comparable with that which it would do if made of (Becquerel's) radium.

Indeed, radioactivity is the driving force behind the Earth's heat, and 'atomic energy', in the form of fusion, powers the sun, not gravitational condensation as Kelvin and Helmholtz had assumed. With this and the advent of radioactive dating, the age of the Earth and the Sun unequivocally jumped to the order of 1,000 million years. (Early in Gribbin's book, an unfortunate misprint indicates 1,000 billion years.)

Arthur Eddington took one of the next critical steps when, in his 1920 address at Cardiff, he suggested that the "transmutation" of four protons into helium, through which 0.7 per cent of the proton mass is converted to energy, is the crucial solar reaction. As Gribbin notes, Eddington's idea was received with scepticism because classical arguments showed that the repulsive Coulomb barrier between protons could not be surmounted at the relatively cold temperature of 40×10^6 K, a value Eddington calculated for the solar core. Furthermore, the chemical composition of the Sun was (erroneously) viewed in this period to mirror the Earth's, thereby implying a paucity of solar hydrogen. Gribbin quotes Eddington's defiant retort to his critics:

The helium which we handle must have been put together at some time and some place. We do not argue with the critic who urges that the stars are not hot enough for this process; we tell him to go and find a hotter place.

Eddington was to be proven correct about the fusion of protons into helium. But it is with this oft-quoted riposte that Gribbin misses the opportunity to point out that Eddington's adamance was driven by a serendipitous misconception: the helium that we observe and to which Eddington refers derives not from stellar fusion but from the fusion occurring in the first moments of the Universe. Without this misconception, the discovery of fusion might well have been postponed for several years. (Also, Gribbin mistakenly lauds Eddington, not the forgotten Homer Lane, for the first application of the perfect gas law to stars.)

In the decades following George Gamow's insight in 1928 that penetration of the Coulomb barrier is allowed quantum-mechanically, there was a sequence of remarkable discoveries that led to a detailed picture of both stellar fusion processes and stellar interiors. One important consequence, as re-

counted by Gribbin, was the quantitative prediction of the flux and energy spectrum of fusion-generated solar neutrinos, ghostly particles that were originally postulated by W. Pauli in 1930 to conserve energy and momentum in laboratory β -particle decays. Because of their weak interaction, neutrinos immediately escape from, and convey information about, the solar core. In contradistinction, fusion energy deposited in the solar core from γ rays and from energetic charged particles requires some 10 million years to reach the Sun's surface, a period fittingly known as the Kelvin-Helmholtz time. Truly we are warmed today by prehistoric energy. But it is these elusive neutrinos that give rise to the prominent solar conundrum: after 25 years of meticulous measurements and calculations, why is the flux of neutrinos that reach the surface of the Earth between one-third and one-half the number predicted? Are both the Brookhaven and Kamiokande neutrino experiments in serious error? Or, alternatively, are we in a situation akin to Kelvin's, in that our rigorous calculations omit an essential, albeit hitherto unknown, element of physics?

Unabashedly, Gribbin takes the latter perspective and advances the imaginative hypothesis based on WIMPs. In essence, WIMPs enhance the transport of thermal energy from the Sun's core to just outside. This reduces the central temperature from 15×10^6 K to 13.5×10^6 K. Since neutrino production is strongly dependent on temperature (to about the seventh power for detected high-energy neutrinos), this ten per cent reduction in core temperature would halve the flux of neutrinos, thereby reconciling theory and experiment. But herein begins the first delicate balancing act: one cannot just lower the central core temperature without also notably reducing the fusion power, which also is strongly dependent on temperature (to about the fourth power). So to maintain a fixed solar output, the central reduction in power needs to be exactly matched by an increase farther out. Supposedly WIMPs do all this and more, prompting John Bahcall in his book *Neutrino Astrophysics* (Cambridge University Press, 1989) to quip that if one must invent a particle, the WIMP at least has the virtue of utility. Bahcall in fact advocates an alternative solution to the neutrino problem that is based on neutrino oscillations, the so-called Mikheyev-Smirnov-Wolfenstein effect. These oscillations, which would render a large fraction of neutrinos unobservable, have just gained further support from the Soviet-American Gallium Experiment (SAGE).

WIMPs have other strengths, according to Gribbin. For example, they have

also been proposed as candidate particles for the dark-matter in the Universe. In this model, 90 per cent of the mass of the Universe resides in weakly interacting particles, their ethereal presence betrayed only by gravitational effects on clusters of galaxies and gas clouds orbiting galaxies. Notwithstanding, I believe that the WIMP hypothesis will remain unconvincing unless WIMPs are directly detected or a compelling theoretical framework is obtained for their existence. The latter, for example, was what Enrico Fermi did for Pauli's neutrinos some 20 years before their direct detection.

To his credit, Gribbin skillfully weaves into the fabric of this story many important physical concepts and concrete numbers without stifling the book's flow. Overall, blemishes are infrequent, although three warrant comment. First, in his discussion of the WIMP hypothesis and the gravitational condensation model, Gribbin implicitly overemphasizes the importance of black-body radiation pressure in comparison to kinetic pressure (that due to electrons and ions). For *solar-mass* stars, the core kinetic pressure is three orders of magnitude larger. Gribbin apparently gets his emphasis from Eddington who, in *The Internal Constitution of the Stars*, gives a value of the kinetic pressure only one magnitude larger than the radiative pressure. Second, although radiative energy transport dominates the first three-quarters of the radial distance of solar-mass stars, the 'resistance' (that is, opacity) to the black-body photons is due to electrons (not protons), electrons in the field of ions, and other mechanisms. And third, the proportion of the Sun's energy flux immediately escaping the solar core via neutrinos amounts to about two per cent, not ten. Surprisingly, perhaps, if it were ten per cent, it is unlikely that any of us would exist. This is because solar evolution and light output would be slightly altered.

Certainly these few points do not detract from the enjoyment and edification provided by the book, and I recommend it. In particular, the recalcitrant solar-neutrino problem, described well by Gribbin, should be viewed as a distinct opportunity for science because it pushes us decisively against the very limits of our knowledge of neutrinos, stellar interiors and particle physics. In his prescient address of 1920, perhaps Eddington overstated the case when he remarked that we know more about the fiery globe overhead than the cooler one underfoot. □

Richard D. Petrasso is at the Plasma Fusion Center, Massachusetts Institute of Technology, 167 Albany Street, Cambridge, Massachusetts 02139, USA.

Between brain and culture

John Morton

Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition. By Merlin Donald. *Harvard University Press*: 1991. Pp. 413. \$27.95, £22.25.

How did the modern human mind evolve from its presymbolic form? Donald proposes that there have been three radical transitions in the evolution of human culture and cognition since the time of our ape-like bipedal ancestors. Two of these transitions are straightforward enough. The first was a change to *Homo erectus* through the emergence of the ability to re-enact events — what Donald calls the “mimetic” skill. The second change, to *Homo sapiens*, is associated with the emergence of speech. It is the final, most recent, nonbiological change that created modern humans.

Like modern apes, our ape-like ancestors were procedurally skilled and had good perceptual and memory abilities. To this was added mimetic skill — the ability to represent knowledge through nonlinguistic voluntary motor acts. These mimetic acts were defined primarily in terms of their representational function and served to create a semantic reference system. The mimetic culture required a degree of social attribution, some teaching skill, and both social coordination and collective knowledge. Many of these skills have in the past been attributed uniquely to language, yet Donald argues not only that the language-less mimetic culture possessed all of them, but that collectively they provided neither a sufficient drive nor a springboard for the invention of language. Language, Donald stresses, “evolved in, and continues to be employed in, a wider cultural context.”

And so to the second important genetically based transition, the development of the human vocal tract, which would not have happened without a radical change in hominid communication; this in turn required the existence of appropriate cognitive skills, driven by the types of mental models that the hominids were creating. The result was a cognitive system that allowed *Homo sapiens* to evolve a complex preliterate culture encompassing causal explanation, prediction and control. Donald calls this the “mythic” culture, a culture which survives in many parts of the world today.

On the other hand, there were three crucial cognitive phenomena that seem to have been underdeveloped in this

oral-mythic culture: graphic invention, external memory and theory construction. These features of modern humans arose through the interaction of cognition with culture. Whereas oral-mythic cultures rely heavily on individual biological memory, modern cultures rely much more on external memory devices, mostly on various symbolic systems ranging from cuneiform and hieroglyphics to alphabetic languages and mathematics. The use of such devices, particularly from a young age, has the effect of reconfiguring the way in which the central nervous system is organized. Thus, for Donald, cognition is the mediator between brain and culture.

Donald argues his case by drawing particularly from evolutionary biology and cultural anthropology but also from cognitive science and neuroscience. His argument accumulates in dribs and drabs rather than growing along foreseen paths. Unusually, he has no axe of his own to grind — this is the only academic book I have read that does not contain any reference to the work of the author. But he does seem to have picked up a few axes belonging to other people on his travels, some of which have become blunt through hard use, and these deviations make some sections of the book unexpectedly hard going. Donald is weak, curiously, on the representation of his theory, having few visual graphics in

the book to represent his mental modelling. And his theory lacks a developmental perspective. Ontogeny must certainly recapitulate phylogeny for Donald's third stage (unless Lamarck is allowed to the party), and there is plenty of relevant work available on cognitive development.

Donald is, of course, the product of his culture as well as his cognition. People are thinking nowadays at several levels about the interaction of genes with environment. There is enormous interest in the role of literacy both in culture and in cognition. And there is work in which the species-typical environment is regarded as being an equal partner to the genotype. So Donald's thesis will undoubtedly find a proper place in the scientific culture. Let me leave you with his final evolutionary challenge — consciousness:

Fear of the homunculus begets irrational behaviour in cognitive scientists. They dread the truth: in a tiny slab of brain there resides a consciousness capable of all we have achieved and experienced; and obviously, on one level, *there is a homunculus*. The homunculus is synonymous with the reflective, conscious mind, and somehow, somewhere in the protean parenchyma of mind, it must reside. □

John Morton is in the MRC Cognitive Development Unit, 17 Gordon Street, London WC1H 0AH, UK.



WHAT is the truth about the nuclear disaster at Chernobyl in 1986? Vladimir Chernousenko, a senior consultant to the Ukrainian Academy of Sciences Commission that was responsible for rectifying the consequences of the accident, claims in *Chernobyl: Insight from the Inside* that between 5,000 and 7,000 people died as a result of the accident and clean-up operations. The official death-toll is 31. But other scientists remain sceptical of his claims (see *Nature* **351**, 4; 1991). Further, a study coordinated by the International Atomic Energy Agency found that the accident did not result in any measurable radiological health effects (*Nature* **331**, 335; 1991). Nonetheless, many Western scientists share Chernousenko's concerns about the safety of scientists still working at Chernobyl (pictured here is the “rudimentary” protective clothing worn by the so-called ‘liquidators’) and the continuing operation of the RBMK-type reactor used there. Published by Springer, price DM68, £24.50.

Plumbing the depths

Peter J. Smith

Water Baby: The Story of Alvin. By Victoria A. Kaharl. Oxford University Press: 1991. Pp.356. £15, \$21.95.

Fire Under the Sea: The Discovery of the Most Extraordinary Environment on Earth — Volcanic Hot Springs on the Ocean Floor. By Joseph Cone. William Morrow: 1991. Pp.287. \$25.

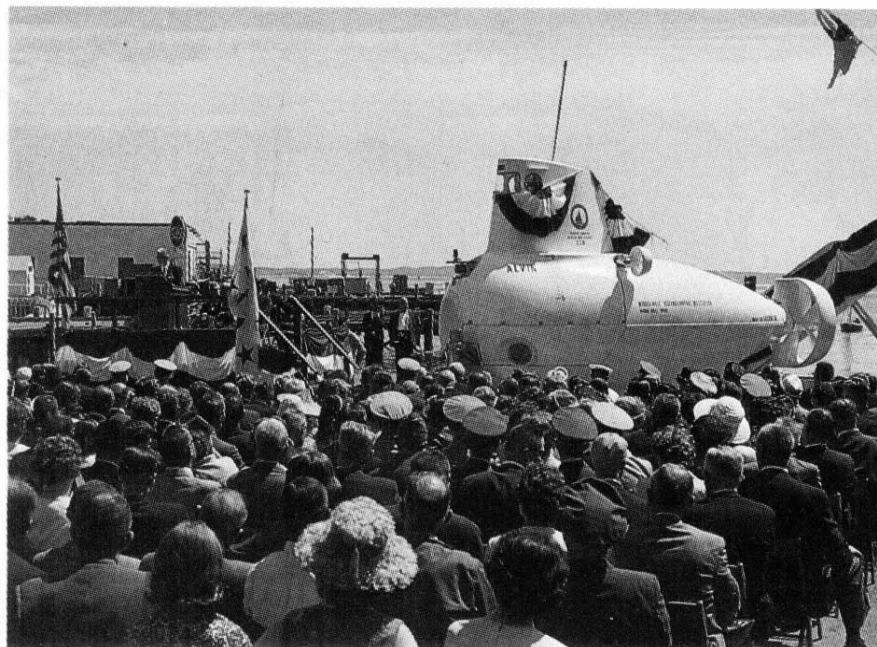
WHEN the custodians of marine nomenclature came snooping round, they were told that the vessel's name was a corruption of that of its conceptual father. In fact, it was named after a cartoon chipmunk popular in the 1950s. But there the joke ended. The deep-diving submarine *Alvin* has proved to be one of the great technological and scientific achievements of the post-war era, even though it is a wonder that it ever got built at all.

When, at a meeting of oceanographers in 1956, Allyn Vine proposed direct exploration of the deep-ocean floor (average depth about 3,900 metres), many of those present did not take him seriously. Most oceanographers (especially those at the Scripps Institution of Oceanography) could not see the point and many never did, the US Navy had no interest in anything below 750 metres, the Bureau of Ships (builder of all US Navy vessels) did not fancy having its authority usurped, and in any case it was by no means clear that the technology was up to it. But a small group of enthusiasts based at the Woods Hole Oceanographic Institution pressed on with faith, technical inventiveness, creative accountancy, some much rarer creative bureaucracy and a little chicanery. There were elements of Ealing comedy about it, not least because the vessel was designed and built by General Mills, the makers of 'Wheaties' and 'Cheeri-oats' — but it worked. *Alvin* was commissioned on 5 June 1964.

The rest is history, but not entirely scientific history. *Alvin* revealed the nonphotosynthetic animal communities around deep hydrothermal vents in the Pacific Ocean and elsewhere, examined the vents' white and black 'smokers' and their associated mineral deposits, collected geological specimens, carried out geophysical observations, photographed phenomena on the ocean floor that had never before been seen, obtained deep-water samples, and discovered modern lithoherms. But it also located the H-bomb that fell into the sea off Spain in 1966, monitored underwater military 'listening posts', examined the state of

sea-bottom nuclear dumps, and detected and explored the *Titanic*. The US Navy soon changed its mind.

Victoria A. Kaharl has produced a gripping account not only of *Alvin*'s birth, life and work, but also of the underside: the perpetual hustling for funds, the arguments between can-do and here-are-ten-reasons-why-you-can't-do-it bureaucrats, the sheer human nastiness that arises when scientific egos compete for priority, the continual need to



Launching a star, *Alvin*, Woods Hole June 1964.

keep undercapitalized technology going with sealing-wax, and the sexual harassment of female oceanographers at sea. The best compliment I can pay Kaharl is to report that one night while reading the book I forgot time and coffee, and suddenly found it was 5.00 a.m. The Woods Hole Oceanographic Institution should also be applauded for appointing a science-writer-in-residence in the first place.

A curious sense of dislocation comes from reading Cone after Kaharl, for he covers part of the same *Alvin* story (that of the investigation of hydrothermal vents) but largely from the point of view of different scientists and different ocean ridges. It is a bit like going to the theatre to see familiar faces and finding understudies in charge, although the effect would no doubt have been the same had the books been read in reverse order. Nor, despite the restricted scope, does one learn much more from Cone about hydrothermal vents, because — and this is a positive point — he provides a much more comprehensive scientific background to the study of vents: how they fit into the global-tectonic scheme of things, and so on. Kaharl, by contrast, is strong on *Alvin*'s technology and operations.

So take your pick: *Alvin* as chief character or the vent story in its wider scientific context. Both approaches have their merits. Both books, it must also be said, take the modern US purplish approach, seeking to impart a sense of authenticity by the use of much extraneous detail and what purports to be direct speech from anything up to 25 years ago. Such literary devices are a matter of taste; and who can complain if they 'bring science alive' to hordes of

nonscientists? Nevertheless, puritans will be uncomfortable with direct speech in historical contexts. How do they *know* what words were used? They do not, of course. I was hoping to catch out Kaharl and Cone by finding them quoting the same supposedly direct speech formulated in different words, but if they did it I missed it. □

Peter J. Smith is in the Department of Earth Sciences, The Open University, Milton Keynes MK7 6AA, UK.

New-year editions

■ Published today by The Johns Hopkins University Press is the revised edition of Jurg Ott's highly acclaimed *Analysis of Human Genetic Linkage*. Price \$47.50, £30.

■ The second edition of *The Structure of Nematodes* by Alan F. Bird and Jean Bird, first published 20 years ago, is issued from Academic Press. The book reviews the structures of all types of nematodes, both free living and parasitic, and now includes a chapter on nematode pathology. Accompanied by 100 photographs, price \$65.

■ Oxford University Press have published the second edition of Wolfgang Rindler's superb *Introduction to Special Relativity*. The book is more conceptually and mathematically than experimentally oriented. Price £30, \$59 (hbk); £15, \$29.95 (pbk).

Protein folding in the cell

Mary-Jane Gething & Joseph Sambrook

In the cell, as *in vitro*, the final conformation of a protein is determined by its amino-acid sequence. But whereas some isolated proteins can be denatured and refolded *in vitro* in the absence of other macromolecular cellular components, folding and assembly of polypeptides *in vivo* involves other proteins, many of which belong to families that have been highly conserved during evolution.

UNTIL recently, scientists using biophysical techniques to study refolding of polypeptides *in vitro* had little need for communication with those investigating the biosynthesis and maturation of proteins within cells. But the realization that the attainment of correct tertiary and quaternary structure is an important determinant of efficient intracellular protein transport¹⁻³ led to the development of techniques to analyse the early stages of protein folding *in vivo*⁴. These studies have shown that in cells families of abundant proteins modulate and promote protein folding, assembly and disassembly, and facilitate the degradation of misfolded polypeptides. Here we review current knowledge of these proteins and discuss current theories of the mechanisms by which they function.

Protein folding *in vitro* and *in vivo*

Although the three-dimensional structures of several hundred proteins are now known in great detail^{5,6}, the pathways by which such polypeptides attain their native configurations remain substantially undefined. Anfinsen's classic experiments on the refolding of ribonuclease *in vitro*^{7,8} established that all the information required to determine the final conformation of a protein can reside in the polypeptide chain itself: the denatured enzyme can refold into its native conformation in the absence of other proteins. Similar results have since been obtained with several other small, single-domain polypeptides, and with a few larger, more complex proteins (reviewed in refs 9-13). Such studies suggest that refolding *in vitro* may be initiated by (1) collapse of hydrophobic regions into the interior of the molecule, (2) formation of stable secondary structures that provide a framework for subsequent folding, and (3) formation of covalent interactions, such as disulphide bonds, that stabilize the polypeptide in particular conformations. Evidence has been obtained in support of each of these mechanisms and it is likely that all three may operate in conjunction during the early stages of refolding. Subsequent folding seems to occur through a limited number of pathways involving distinct intermediates ('molten globules'^{14,15} or 'compact intermediates'¹³) that have significant secondary structure and a compact form but lack a well-defined tertiary structure and expose more hydrophobic surface than fully folded molecules. These intermediates seem to be in rapid equilibrium with the fully denatured state and are only slowly converted to the native state. Thus the rate-limiting step in the refolding process frequently occurs at a very late stage, just before the protein adopts its final, native conformation.

Despite their value in defining the types of intramolecular interactions that drive polypeptide folding, *in vitro* experiments do not accurately reflect the process of folding of nascent proteins in the interior of a cell. Refolding *in vitro* is frequently very inefficient in comparison to folding *in vivo*, and often requires protein concentrations and physicochemical conditions very different from those occurring intracellularly. Furthermore, although refolding experiments involve the whole polypeptide chain, the opportunity exists *in vivo* for folding to commence as soon as the N-terminal portion of the nascent chain emerges from the ribosome (or from the lipid bilayer following membrane translocation). Finally, many proteins exist in cells as homo- or hetero-oligomeric complexes that in some cases are assembled

before folding of the individual polypeptide chains is complete. The probability is slight that such complexes could form *in vitro* at the low subunit concentrations occurring intracellularly¹⁶; the probability is even lower that they could form before individual chains have become irreversibly misfolded.

To investigate protein folding *in vivo* it was first necessary to devise assays for polypeptide conformation that do not depend on obtaining large quantities of partially folded proteins that are sufficiently pure for physicochemical measurements. The first such assays analysed the formation of disulphide bonds during the folding of biosynthetically labelled secretory proteins and showed that these bonds can form even before synthesis of a polypeptide is completed, and that disulphide bond formation occurs *in vivo* at a significantly faster rate than can be achieved under the most favourable conditions *in vitro*¹⁷. Subsequently, additional *in vivo* folding assays were developed that use conformation-specific antibodies, protease sensitivity, or sucrose density gradient centrifugation to probe the tertiary and quaternary structure of radiolabelled proteins (reviewed in ref. 4). Studies using these techniques showed that although individual domains of a nascent polypeptide may fold very rapidly, acquisition of the final native structure of the whole protein can proceed comparatively slowly. Furthermore, partially folded intermediates whose structures would seem unlikely to be stable in the presumably aqueous environment of the cell's interior can lie dormant for many minutes to many hours before folding is completed^{3,18,19}. Despite such extended pauses in the assembly process, folding *in vivo* of wild-type proteins is usually highly efficient with >95% of the newly synthesized polypeptides eventually attaining their native three-dimensional structures^{3,20}. Polypeptide misfolding and aggregation, frequently a major problem during refolding *in vitro*²¹, rarely occurs *in vivo* except with mutant proteins or during synthesis at elevated temperatures. Finally, partially folded polypeptides can frequently be isolated as complexes with specific cellular proteins, notably members of stress protein families^{3,22-26}.

The rest of this article summarizes the results of studies that have revealed the existence of at least two classes of proteins involved in polypeptide folding in cells. The first class includes conventional enzymes that catalyse specific isomerization steps that may otherwise limit the rate of folding of some proteins, whereas a second class of 'chaperones' stabilize unfolded or partially folded structures and prevent the formation of inappropriate intra- or interchain interactions. Some members of this second class also interact with apparently native protein molecules to promote rearrangement of protein-protein interactions in oligomeric structures.

Enzymes involved in protein folding

In vitro, two rate-determining steps involving isomerization of covalent bonds can be catalysed by purified cellular enzymes. Protein disulphide isomerase (PDI) catalyses thiol/disulphide interchange reactions and, depending on the nature of the polypeptide substrate and the imposed redox potential, promotes protein disulphide formation, isomerization or reduction^{27,28}. PDI does not determine the polypeptide's folding pathway, but rather facilitates formation of the correct set of disulphide bonds

by promoting rapid reshuffling of incorrect disulphide pairings. Proteins with peptidyl prolyl *cis-trans* isomerase (PPIase) activity catalyse the otherwise slow isomerization of X-P peptide bonds (where X is any amino acid and P is proline in single-letter amino-acid code) and can accelerate the refolding of proline-containing polypeptides *in vitro*^{10,29,30} and *in vivo*³¹.

PDI and the thioredoxin-like proteins

Several lines of evidence suggest that PDI activity is required for folding of nascent polypeptides in the endoplasmic reticulum (ER) of eukaryotic cells. Thus the abundance of the enzyme in the ER of different cell types correlates with the level of secretory protein synthesis¹⁷ and *in vivo* chemical crosslinking studies demonstrate a specific association between the enzyme and newly synthesized immunoglobulin chains in the ER (ref. 32). Finally, reintroduction of purified PDI into microsomes evacuated of their luminal content by alkali or detergent treatment restores cotranslational formation of disulphide bonds in proteins synthesized using a cell-free system³³. PDI is an essential protein in yeast^{34,35}, but the biological process whose disruption leads to lethality has not yet been defined.

Mammalian PDI is a dimer of identical subunits of relative molecular mass 57,000 (M_r 57K), each of which contains duplications of domains showing strong homology to thioredoxin³⁶, a small redox protein present in all classes of organisms from bacteria to higher eukaryotes³⁷. Computer modelling studies based on the known three-dimensional structure of *Escherichia coli* thioredoxin indicate that a functional PDI dimer contains four thioredoxin-like domains each having a dithiol/disulphide active site located on a prominent loop at the surface of the molecule³⁸.

PDI's role in the ER may not be limited to disulphide isomerization²⁸. First, many types of mammalian cells contain, in addition to substantial amounts of homodimeric PDI, the enzyme prolyl-4-hydroxylase which causes extensive modification of proline residues in nascent collagen molecules. This enzyme, an $\alpha_2\beta_2$ tetramer whose β -chain dimers are identical to PDI, also has PDI activity (reviewed in ref. 27). The 64K α -chains form the binding site for the peptide substrate to be hydroxylated, but it has not yet been determined whether the dithiol/disulphide active sites of the PDI/ β -subunits are directly involved in the hydroxylation reaction. Second, the 'glycosylation site binding protein' (GSBP) component identified using a glycosylation site photoaffinity probe³⁹ is identical to PDI (ref. 34). However, involvement of this enzyme during N-linked glycosylation in the ER is not proved, as depletion of PDI from microsomes does not affect their capacity to support oligosaccharide addition to nascent polypeptides⁴⁰. Finally, PDI has also been identified as a component of the microsomal triglyceride transfer protein complex⁴¹.

Recent studies have identified other ER proteins containing thioredoxin homology units. ERp59, ERp61 and ERp72, three members of a set of proteins whose synthesis is induced at onset of immunoglobulin secretion in murine B cells⁴², have been characterized by complementary DNA cloning and sequencing. ERp59 is identical to PDI (ref. 43). Although ERp61 contains two thioredoxin-like domains found in the same relative positions as in the PDI molecule, the rest of its sequence is unrelated to that of PDI (R. Mazzarella and M. Green, personal communication). ERp72 contains three thioredoxin homology units, two of which are spaced as in PDI, embedded in otherwise unrelated sequences⁴³. Neither ERp61 nor ERp72 have PDI activity *in vitro*. Finally, an essential gene, *EUG1*, encoding another PDI-related ER protein containing thioredoxin homology units but lacking PDI activity, has been identified in *Saccharomyces cerevisiae* (C. Tachibana and T. Stevens, personal communication). Whether the unknown functions of these proteins use the redox activity of their thioredoxin-related structural units is still in question. In this context it is of interest that thioredoxin itself is required for assembly of filamentous phages in *E. coli*, playing

a part that does not involve its redox activity as site-specific mutation of either or both of the active site cysteines does not alter the ability of the mutant proteins to support phage assembly (for review, see ref. 44). Thioredoxin is thought to confer processivity on the reaction that leads to the displacement of the intracellular phage protein pV from single-stranded phage DNA and to its replacement at the membrane by the major coat protein, pVIII.

Thus the ER houses an extended family of enzymes (Table 1) that may use thioredoxin-like domains containing dithiol/disulphide active sites to carry out various functions in the co- and post-translational modification of secretory proteins. Whether these proteins may also have roles in protein assembly that do not involve redox activity, and whether there may be, in addition to thioredoxin itself, other members of this family located in other compartments of the cell remains to be determined.

Peptidyl prolyl *cis-trans* isomerases

Proteins with PPIase activity are highly abundant and widely distributed, being found in virtually all tissues and organisms, from bacteria to mammals (reviewed in ref. 10). Those proteins characterized so far fall into two structurally unrelated families (Table 1), which are named after the clinically important immunosuppressive agents that inhibit their isomerase activity. Thus the eukaryotic cyclophilin proteins bind cyclosporin A (CsA) with high affinity, whereas the FK506-binding proteins bind the structurally distinct compounds FK506 and rapamycin (reviewed in ref. 45). Both CsA and FK506 mediate their immunosuppressive action by preventing the transcription of genes involved in activation of T lymphocytes, whereas rapamycin potentially inhibits the response of T cells to the lymphokine IL-2 (ref. 45). These immunosuppressive drugs do not act through inhibition of the PPIase activity of T cells as they are effective at concentrations far below those of the PPIase enzymes, and they inhibit distinct signalling pathways. Rather it seems that cyclophilin and FKBP bind the drugs, which are cyclic peptides, and present them (to as yet undefined targets) in a bioactive conformation that may require a *cis-trans* isomerization around one of their peptide bonds^{46,47}.

Although most of the PPIase activity in cells is found in the cytosol, family members are located in different cellular compartments. Thus cyclophilin-like proteins are also present in the periplasmic space of *E. coli* cells^{48,49}, and in the mitochondrial matrix of *Neurospora crassa*⁵⁰. Furthermore, the nucleotide sequences of cyclophilin-related genes cloned from *S. cerevisiae*⁵¹, *Drosophila melanogaster*^{52,53} and vertebrate cells⁵⁴⁻⁵⁷ and of the FKBP-related gene from human cells⁵⁸ each encode a stretch of N-terminal amino acids whose sequences are compatible with function as signal peptides for translocation into the lumen of ER. Following their translocation into the ER, some cyclophilin-related proteins may also be transported along the exocytic pathway and secreted into the extracellular medium^{55,56}. Consistent with this diversity of localization are the findings that multiple genes encoding cyclophilin- and FKBP-related proteins are present in mammalian cells and lower eukaryotes⁵⁸⁻⁶¹, although in *N. crassa* a single gene encodes both cytosolic and mitochondrial cyclophilins⁵⁰. Genetic studies in lower eukaryotes demonstrated that the cyclophilin and FKBP gene products that mediate sensitivity to CsA or FK506 and rapamycin are not essential for cell viability either individually⁶¹⁻⁶³ or in combination⁶³, either because of the presence in cells of additional proteins having PPIase activity or because this activity is not required for cell survival.

Despite our partial understanding of their adventitious role in the suppression of T-lymphocyte function⁴⁵, the real *in vivo* role and physiological substrates of enzymes with PPIase activity remain to be established. It seems likely that these highly abundant and widely distributed proteins normally act as 'conformases' (ref. 10) catalysing slow steps in the initial folding and/or rearrangement of protein structures. Initial evidence in support

TABLE 1 Enzymes and chaperones that may be involved in protein folding and assembly in cells

Organism/organelle		Enzymes			Chaperones		
	Protein family	PDI	Cyclophilin PPIase	FKBP PPIase	Hsp60 (Chaperonin-60)	Hsp70 (Stress-70)	Hsp90 (Stress-90)
<i>E. coli</i>							
Cytosol		Thioredoxin	PPIase b		GroEL	DnaK	HtpG (C62.5)
Periplasm			PPIase a (Rotamase)				
Yeast							
Cytosol			Cph1p (Cpr1p)	Fkb1p (Fkr1p) (Rbp1p)		Ssa1–4p	Hsp83 Hsc83
ER		PDI Eug1p	yCyPB			Kar2p (BiP)	
Mitochondria					Hsp60 (Mif4p)	Ssc1p	
Drosophila							
Cytosol			CyP			Hsp68 Hsp70 Hsc1,2,4	Hsp83
ER		PDI	NinaA				
Mammals							
Cytosol			Cyclophilin (PPIase) (CyPA)	FKBP		Hsp70 (p73) Hsc70 (p72) (CUATPase) (Prp73)	Hsp90 (Hsp83) (Hsp87)
ER		PDI (ERp59) GSBP ERp72 ERp61	CyPB(rCyPLP)			BiP (Grp78)	Grp94 (ERp99) (endoplasmin)
Mitochondria					Hsp60 (Hsp58)	Hsp70 (Grp75)	
Plants							
Cytosol							
ER		PDI				b70 (BiP)	
Chloroplasts					RuSBP		

Six protein families have been identified whose members include enzymes or chaperones proposed to be involved in folding, assembly, rearrangement or degradation of proteins in cells. Members that have been characterized to date from a variety of different organisms are shown. Alternative names are shown in parentheses. For references see the text.

of this hypothesis came from the discovery that the *D. melanogaster ninaA* gene product, an eye-specific cyclophilin-related membrane protein, is required for the folding and/or stability of rhodopsins 1 and 2 (refs 52, 53, 64). Treatment of chick embryo fibroblasts with cyclosporin A delays the folding of the triple helix of type I collagen, indicating a physiological role for cyclophilin PPIase in folding in the ER (ref. 31).

The mechanism by which the two classes of PPIases catalyse rotation around specific peptide bonds also remains to be determined. Human cyclophilin and FKBP display dramatic differences in substrate specificity. Although cyclophilin has a broad specificity and does not discriminate between P₁ amino-acid residues, FKBP has a narrow specificity with a preference for hydrophobic residues at the P₁ position⁶⁵. The reason for this preference became clear with the determination of the three-dimensional structure of FKBP (refs 47, 66), as the folding topology provides a large cavity, lined with conserved aromatic residues, that serves as the active site and drug binding pocket. Mechanistic studies (reviewed in ref. 45) suggest that both cyclophilin and FKBP catalyse the interconversion of *cis*- and *trans*-rotamers of peptide substrates by noncovalent stabilization of a twisted amide transition state.

Protein chaperones

A number of other cellular proteins, now collectively known as chaperones⁶⁷, function *in vivo* not as catalysts of secondary structure formation, but rather to recognize and stabilize partially folded intermediates during polypeptide folding, assembly and disassembly. The majority of the currently identified chaperones belong to three highly conserved protein families, whose members are widely distributed from prokaryotes to

plants and mammals (Table 1). In eukaryotic cells, different family members are found in different cellular compartments and organelles. As their names imply, proteins of the hsp70, hsp90 and chaperonin (groEL/hsp60) families first came to attention because of their specific induction during the cellular response of all organisms to heat shock (reviewed in refs 68–71). Nevertheless, the majority of the family members are expressed constitutively and abundantly in the absence of any stress, and genetic studies show that many of these proteins are essential for cell viability under normal conditions of growth^{68,72,73}. Many hsp family members, including those that do not respond significantly to heat shock, are induced under a variety of other stress conditions^{68,74,75} whose common denominator may be the accumulation of unfolded or misfolded proteins in cells^{76–79}. To reduce confusion arising from the use of the term 'hsp' to describe all family members regardless of their response to heat shock, we will use the terms 'stress-70' and 'stress-90' for members of the 70K and 90K families, respectively, a nomenclature similar to that suggested by Craig^{68,80} for the 70K family of *S. cerevisiae*. The 60K (groEL/hsp60) family is currently described by the term 'chaperonin-60'^{81,82}.

Chaperones seem to serve many functions that stem from their ability to recognize and modulate the state of folding of polypeptides within cells (Table 2). Thus members of the stress-70 family have been implicated in the stabilization or generation of unfolded protein precursors before assembly in the cytosol⁸³ or translocation into organelles including the ER and mitochondria^{84–86}, in stabilization of newly translocated polypeptides before folding and assembly^{3,23,87–90}, in rearrangement of protein oligomers^{72,91,92}, in dissolution of protein aggregates^{87,93}, and in the degradation of rapidly turned-over

cytosolic proteins^{94,95}. Stress-90 proteins seem to stabilize a variety of target proteins in an inactive or unassembled state⁶⁸. Chaperonin-60 proteins bind unfolded precursors before export of secretory proteins^{24,26} or assembly of protein oligomers⁹⁶ in *E. coli*, and help in folding and assembly of polypeptides translocated into chloroplasts or mitochondria in eukaryotic cells⁹⁷⁻⁹⁹. In addition, unrelated cellular proteins that are not members of any of these stress protein families have been implicated in protein in cells. Nucleoplasmin, for which Laskey^{100,101} coined the term chaperone in 1978, binds to histones and facilitates ordered nucleosome assembly in the nucleus. In the ER, the T-cell receptor-associated protein, TRAP (ref. 102) or p28 (ref. 103), is noncovalently associated with the newly synthesized CD3 chains until they assemble with other subunits of the receptor, whereas an 88K protein transiently associates with newly synthesized major histocompatibility complex (MHC) class I heavy chains¹⁰⁴. In *E. coli*, SecB and trigger factor also bind unfolded secretory precursors before their export across the plasma membrane^{26,105-107}. PapD has been proposed to act as a chaperone in the periplasmic space to enhance the folding and assembly of components of P pili^{108,109}, and 'scaffolding proteins' encoded by bacteriophages promote the assembly of phage coats although they are not finally incorporated into the viral particles¹¹⁰. With the realization that polypeptide folding in the cell frequently requires the assistance of chaperones, some with a broad target specificity and some dedicated to assembly of particular macromolecules, it seems inevitable that many more such proteins will be identified in the coming months and years.

The stress-70 protein family

The role of stress-70 proteins during the heat-shock response has been studied extensively for many years (for recent reviews see refs 68-71), but only recently has their importance in normal cellular processes such as protein folding, assembly, disassembly and degradation (Fig. 1) become widely appreciated^{87,95,111,112}. Stress-70 family members implicated in such processes include in *E. coli*, DnaK; in yeast the cytosolic proteins Ssa1p and Ssa2p, the ER protein Kar2p and the mitochondrial protein Ssc1p; and in mammalian cells the cytosolic proteins hsp70 (p72), hsc70 (p73, clathrin uncoating ATPase) and prp73 (peptide recognition protein 73), and the ER protein BiP (also known as grp78). Although in no case is the interaction between an individual stress-70 protein and its target polypeptide understood in molecular detail, the available evidence reveals several common features that point to a conserved mechanism for the action of these ubiquitous proteins. These features include (1) differential recognition of target polypeptides and modulation of their conformation or state of assembly, (2) involvement of ATP binding and/or hydrolysis, (3) a requirement for other heat shock proteins or cellular factors, and (4) the induction of synthesis of individual stress-70 family members by the accumulation of unfolded proteins in the appropriate cellular compartment.

***E. coli* DnaK.** DnaK, originally, defined as the product of a host gene required for bacteriophage λ DNA replication in *E. coli*, also plays fundamental roles in normal cellular physiology⁷². Mutations in the *dnaK* gene result in temperature-sensitive growth of *E. coli*, overproduction of other heat-shock proteins even at permissive temperatures, impaired synthesis of DNA and RNA and a generalized defect in proteolysis. Furthermore, the synthesis of DnaK is

increased as a result of the accumulation in cells of unfolded polypeptides¹¹³ and DnaK binds foreign eukaryotic proteins expressed in *E. coli*¹¹⁴. Finally, overproduction of DnaK aids the export of lacZ hybrid proteins across the bacterial inner membrane¹¹⁵. These observations all indicate the involvement of DnaK in modulating many protein-protein interactions *in vivo*. Detailed *in vitro* studies have now illustrated the ability of DnaK to interact with either fully assembled or unfolded polypeptide substrates. During bacteriophage replication, DnaK functions together with two other heat shock proteins, DnaJ and GrpE, to release lambda P protein from the inactive pre-primosomal replication complex^{72,116,117}. DnaK also functions together with DnaJ to activate RepA initiator protein for binding to the origin of replication of plasmid P1 (ref. 118). Hydrolysis of ATP, thought to be catalysed by DnaK, is required during both reactions, and *in vitro* this ATPase activity of DnaK can be stimulated up to 50-fold in the simultaneous presence of DnaJ and GrpE (ref. 119). Finally, DnaK, which often associates with *E. coli* RNA polymerase through many purification steps, can protect the enzyme from heat inactivation *in vitro*⁹³. ATP is not required for the protective effect. DnaK can also reactivate heat-inactivated RNA polymerase by dissolving aggregates formed at high temperatures but this process is absolutely dependent on the hydrolysis of ATP.

Cytosolic stress-70 proteins of eukaryotes. On heat shock of mammalian cells, both the constitutively expressed hsc70 proteins and the heat-induced hsp70 proteins migrate from the cytoplasm to the nucleus where they associate with polypeptides that form an insoluble complex at the increased temperature (reviewed in ref. 111). Subsequently the stress-70 proteins also migrate to the nucleolus and associate with partially assembled prribosomes. Presumably, at elevated temperatures nuclear proteins become partially denatured, exposing hydrophobic regions that tend to interact to form insoluble aggregates. Pelham^{87,120} proposed that by binding to the exposed hydrophobic surfaces, stress-70 proteins could limit such interactions and perhaps promote disaggregation. In cell extracts, the stress-70 proteins could be released from their association with nucleolar proteins by addition of ATP, but not nonhydrolysable ATP analogues¹²⁰.

In *S. cerevisiae*, the constitutively expressed cytosolic hsc70 proteins Ssa1p and Ssa2p are required, together with a separate

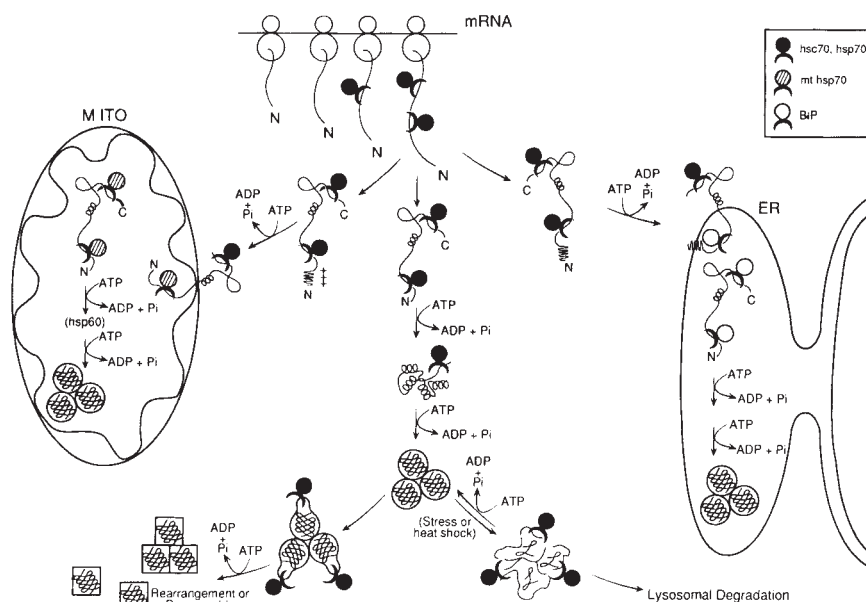


FIG. 1 Illustration of the proposed roles of stress-70 proteins in eukaryotic cells during the folding and membrane translocation of nascent polypeptides, during molecular rearrangements or disassembly, in protection from stress and in protein turnover.

TABLE 2 *In vivo* roles of protein chaperones

	Target polypeptide	Chaperone	Role	Reference
E. coli	Secretory precursors (β -lactamase, proOmpA, prePhoE)	GroEL SecB Trigger Factor DnaK?	Antifolding before translocation	24, 26, 105, 106, 107, 115, 194
	DNA replication complexes	DnaK/DnaJ	Rearrangement of protein complex	72, 116, 117
	Bacteriophage head or tail proteins	GroEL/GroES Bacteriophage scaffolding proteins	Head or tail assembly	96 110
	Foreign proteins	DnaK	Stabilization of unfolded protein?	114
	P pili in periplasmic space	PapD	Pili assembly	118
Photosynthetic bacteria	Rubisco	GroEL/GroES	Oligomer assembly	82, 188, 192, 195, 196
Chloroplasts	Rubisco	RuSBP	Oligomer assembly	81, 191, 192
Mitochondria	Mitochondrial precursors	Hsp70	Completion of translocation Stabilization of prefolded structures in matrix	90, 147 90
	Precursors in matrix	Hsp60	Stabilization of prefolded structures and folding Re-export of precursors to intermembrane space	204–206 204
ER	Nascent secretory proteins	BiP	Completion of translocation Stabilization of prefolded structures in lumen	138 3, 23, 89, 133, 134
	Mutant or foreign proteins	BiP	Stabilization of unfolded structures?	3, 151, 152
	Subunits of T cell receptor MHC Class I heavy chains	TRAP or p28 p88	Receptor assembly Stabilization of newly synthesized heavy chains?	102, 103 104
	Plant storage proteins	b70 (BiP)	Stabilization of newly synthesized polypeptides?	127, 132
Cytosol	Nascent polypeptides	Hsc70 (Hsp70?)	Stabilization of prefolded structures?	83
	Mitochondrial and secretory precursors	Hsc70 (Hsp70?)	Antifolding before translocation	84–86
	Clathrin-coated vesicles	Hsc70 (clathrin uncoating ATPase)	Binds exposed loop of clathrin light chain to promote uncoating	91, 92
	Aged? proteins	Hsc70 (Prp73)	Targeting to lysosomes for degradation	94, 95, 124
	Steroid receptors	Hsp90	Stabilizes inactive form of receptor	178–180
	Retroviral transforming proteins	Hsp90	Stabilizes inactive form of protein during transit to plasma membrane	176, 177
	Actin, tubulin	Hsp90	Stabilizes protein subunits?	68
Nucleus	Preribosomes	Hsp70/Hsc70	Protection of heat denatured proteins	111
	Histones	Nucleoplasmin	Nucleosome assembly	100, 101

N-ethylmaleimide(NEM)-sensitive cytosolic factor, for membrane translocation of secretory and mitochondrial precursors^{84–86}. Consistent with this role is that the synthesis of cytosolic stress-70 proteins is induced in yeast cells by the accumulation of secretory precursors in the cytoplasm⁷⁹. Because in cell-free experiments the need for the stress-70 proteins can be eliminated by urea-mediated unfolding of precursors^{84,86}, the hsc70 proteins are thought to promote a translocation competent state by relaxing the tertiary structure of the polypeptides or by dissolving aggregates of untranslocated precursors^{84,85}. *In vivo*, the hsc70 proteins may bind to nascent secretory precursors before they fold, maintaining them in a translocation-competent state before membrane penetration. Both Ssa1p and Ssa2p bind ATP with high affinity^{84,85}, but the involvement of ATP during their interaction with secretory proteins has not been established. *In vitro* experiments indicate that

mammalian hsc70 can function in the same manner as Ssa1p and Ssa2p to stimulate import of M13 procoat into mammalian ER microsomes¹²¹. Hsc70 also interacts transiently with newly synthesized cytosolic proteins⁸³, presumably to facilitate their proper folding and subsequent assembly in the cytoplasm.

In vivo the targets of the cytosolic stress-70 proteins are not limited to damaged proteins or nascent polypeptides as hsc70, in the form of clathrin-uncoating ATPase^{122,123} also promotes the disassembly of clathrin cages by displacing triskelions from the clathrin lattice in a process that requires ATP hydrolysis and the presence of clathrin light chains (reviewed in ref. 91). Transient changes in Ca^{2+} and/or K^{+} concentrations in the cytosol surrounding a newly invaginated clathrin-coated vesicle apparently cause exposure of a stretch of amino acids (residues 47–71) in LC_a light chains that comprise a binding site for hsc70 (ref. 92). The interaction between the LC_a peptide and hsc70,

which *in vitro* both alters the conformation of the hsc70 molecule and stimulates its ATPase activity, then initiates the uncoating process which proceeds in a cooperative manner. That the hsc70 binding site on the LC_a molecule is cryptic under the ionic conditions normally present in the cytosol explains why the large cellular excess of hsc70 over clathrin does not lead to a permanent state of clathrin disassembly⁹².

Finally, prp73, a stress-70 family member that is almost certainly identical to hsc70, is involved in the lysosomal degradation of intracellular proteins⁹⁴. Prp73 binds peptide sequences (KFERQ and related sequences^{95,124}) that target intracellular proteins for lysosomal degradation in response to serum withdrawal. When lysosomal uptake and degradation of protein substrates is reconstituted *in vitro*, prp73 stimulates degradation in an ATP-dependent manner⁹⁴. Serum starvation causes induction of prp73, which presumably alters the conformation of KFERQ-containing proteins so that they can be translocated across the lysosomal membrane.

Stress-70 in the endoplasmic reticulum. The ER of eukaryotic cells contains a roughly 78K member of the stress-70 protein family^{79,125-127}, now named BiP^{23,88}. In mammalian cells, this protein was originally described independently as the immunoglobulin heavy chain binding protein^{22,128}, and the glucose-regulated protein, grp78 (ref. 129). In yeast cells, BiP is the product of the *KAR2* gene^{79,126}, one of a class of genes involved in nuclear fusion following mating of yeast cells^{130,131}. Under normal growth conditions, BiP is synthesized constitutively and abundantly and comprises about 5% of the luminal content of the ER of mammalian cells. Its synthesis can be further induced by the accumulation of secretory precursors⁷⁹ or mutant proteins in the ER (refs 78, 132), or by a number of different stress conditions⁷⁵ that also lead to the accumulation in the ER of unfolded polypeptides⁷⁸. BiP associates transiently with a variety of nascent wild-type exocytotic proteins^{3,23,89,133,134} and more permanently with misfolded or unassembled proteins whose transport from the ER is blocked^{3,23,135}. Complexes between BiP and nascent secretory proteins, isolated from extracts of mammalian cells, can be dissociated *in vitro* by the addition of ATP, but not of nonhydrolyzable analogues or ADP (ref. 125). BiP, an essential protein in yeast^{79,126,136}, is therefore thought to have a role in the folding and assembly of newly synthesized proteins in the ER lumen^{3,23,87,88}. Initial suggestions that BiP might recognize and retain unfolded proteins in the ER^{3,23,87,137}, or target misfolded proteins for destruction¹³⁵ are no longer tenable. Thus, although unassembled immunoglobulin heavy chains are secreted if the BiP-binding domain is deleted¹³⁷, a truncated form of influenza haemagglutinin that does not bind BiP is neither secreted, nor degraded more slowly, than other transport-defective haemagglutinin mutants that do bind to BiP (M. Segal, J.F.S. and M.-J.G., unpublished results). In addition to modulating protein folding in the ER lumen, BiP may be directly or indirectly involved in translocation of precursors across the ER membrane. Yeast cells expressing a temperature-sensitive *kar2* mutant accumulate secretory precursors in the cytosol at the nonpermissive temperature¹³⁸. In addition, *KAR2* interacts genetically with *SEC63* (J. Vogel and M. Rose, personal communication), a yeast gene that encodes a transmembrane protein required for ER translocation of secretory precursors¹³⁹. Sec63p spans the ER membrane and extends into the lumen a domain containing sequences homologous to DnaJ (ref. 140), another heat-shock protein frequently required for DnaK's function in *E. coli*^{72,118,119}. Whether cooperation with DnaJ-related proteins is required for the chaperone activity of all stress-70 family members remains to be determined.

Stress-70 in the mitochondrial matrix. Stress-70 proteins have been identified in the mitochondria of a number of organisms including *S. cerevisiae*¹⁴¹, *Euglena gracilis*¹⁴², *Trypanosoma cruzi*¹⁴³ and mammals^{144,145}. The yeast protein (Ssc1p) has been localized to the mitochondrial matrix⁹⁰ where it performs an essential function, as disruption of the *SSC1* gene is lethal¹⁴⁶.

The amino-acid sequence of Ssc1p is more closely related to DnaK than are those of the other eukaryotic stress-70 proteins¹⁴¹, consistent with the presumed endosymbiotic origin of mitochondria. Studies using a temperature-sensitive *ssc1* yeast mutant⁹⁰ demonstrate dual functions for Ssc1p that parallel those suggested for the ER-located BiP protein. These are involvement in translocation of precursor proteins through the lipid bilayer at mitochondrial contact sites, and involvement in folding of the imported polypeptides in the mitochondrial matrix. The translocation defect can be circumvented *in vitro* by artificially denaturing the precursor molecules, allowing investigation of events in the matrix of isolated mitochondria containing mutant forms of Ssc1p. Interestingly, the imported precursors remain in an unfolded state and can be isolated in physical association with the mutant Ssc1p protein. On the basis of these observations, Kang *et al.*⁹⁰ proposed that Ssc1p binds the precursor polypeptide as it emerges on the matrix side of the translocation apparatus in contact sites⁹⁷, supporting the continuation of translocation by 'pulling' the precursor into the matrix space. Subsequently, Ssc1p would maintain the imported precursor in an unfolded state until it is released, possibly in an ATP-dependent step, for subsequent folding catalysed by other components in the mitochondrial matrix (see Fig. 2 and discussion below of the role of the chaperonin hsp60). The isolation by crosslinking of complexes containing a partially translocated precursor, a mitochondrial outer membrane protein (ISP42) and Ssc1p provides direct evidence that Ssc1p can interact with polypeptides before the chains have been completely imported into the mitochondrial matrix¹⁴⁷.

Role of ATP binding and hydrolysis by stress-70 proteins. All stress-70 family members bind ATP and a number of them, including DnaK (ref. 16), hsc70 (ref. 122) and BiP (ref. 148), have weak ATPase activities that can be elicited by appropriate protein substrates and by some but not all synthetic peptides^{92,148}. Adenine nucleotide binding apparently causes conformational changes in stress-70 proteins that result in altered sensitivity to proteases¹⁴⁹, or in alteration of their oligomeric state¹⁵⁰. ATP and ADP differ in their effects. ATP protects a roughly 60K fragment of BiP while ADP protects a roughly 45K fragment; ATP stabilizes the monomeric form of hsc70 whereas ADP stabilizes the dimer. These consequences of ATP binding do not require hydrolysis as the nonhydrolyzable analogue ATPγS can substitute. By contrast, addition of ATP, but not of nonhydrolyzable analogues, to cell extracts causes dissolution of complexes between stress-70 proteins and their polypeptide substrates, including DnaK and bacteriophage λP protein¹¹⁷, hsp70 and heat-shocked nuclei¹²⁰, hsc70 and mutant forms of the cellular p53 protein¹¹⁴, and BiP and immunoglobulin heavy chains¹²⁵. These observations led Pelham^{87,120} to propose that ATP hydrolysis causes conformational changes in stress-70 proteins that are transmitted to the substrates, promoting their folding or weakening their interactions with other polypeptides. However, it is currently believed that binding of stress 70 proteins may simply stabilize unfolded conformations of their target polypeptides, preventing the formation of inappropriate intra- or intermolecular interactions. Rothman¹⁶ has proposed that ATP hydrolysis, which takes place *in vitro* with a turnover time of about 5 min, provides a timed mechanism of release of the stress-70 protein from its substrate, freeing the polypeptide to continue the folding process. This turnover time could of course be altered *in vivo* by interactions with other cellular components. As discussed earlier, the ATPase activity of DnaK can be stimulated 50-fold in the presence of DnaJ and GrpE proteins¹¹⁹.

Substrate recognition by stress-70 proteins. The molecular basis of substrate recognition by stress-70 proteins remains a matter for speculation. Prp73 binds cytosolic proteins that contain sequences identical or closely related to the consensus pentapeptide KFERQ^{94,95,124}. Other stress-70 family members interact with a variety of polypeptides that do not contain any conserved

motif, suggesting that their 'recognition signals' do not consist of unique linear amino-acid sequences. Despite their broad patterns of polypeptide recognition, these proteins are not interchangeable. Thus clathrin cages elicit the ATPase activity of hsc70 but not that of BiP (ref. 148), and neither DnaK nor BiP can replace prp73 in promoting targeting of RNAase A for lysosomal degradation⁹⁴. Evidence is lacking, however, for evolution of structural features that limit recognition by individual stress-70 proteins to *bona fide* targets, as DnaK will bind eukaryotic cellular or viral proteins when they are expressed in *E. coli*¹¹⁴, and BiP will recognize bacterial enzymes or virally encoded nuclear antigens when they are artificially introduced into the ER lumen by virtue of addition of an N-terminal hydrophobic signal sequence^{151,152}. Nevertheless, target recognition by these proteins is not indiscriminate: many examples exist of authentic secretory proteins that do not seem to associate with BiP. Similarly, high-affinity BiP binding may be confined to specific domains in individual polypeptides. Thus the CH₁ domain of immunoglobulin heavy chain is necessary for stable interaction with BiP (ref. 137), and binding of BiP to influenza haemagglutinin is apparently limited to sequences in the stem domain (M. Segal, J. S. and M.-J.G., unpublished results). In these cases BiP seems to interact with sequences that form subunit interfaces because the CH₁ domain of the immunoglobulin heavy chain is the site for docking of the light chain¹³⁷, whereas the haemagglutinin trimer assembles through cooperative folding of sequences in the stem domains^{3,153}. Some proteins or protein domains may lack appropriate recognition signals and never interact with BiP. Alternatively, BiP may be involved in the folding of all nascent molecules in the ER but some interactions may be too transient to be detected experimentally, possibly because BiP interacts with different segments of polypeptides with a spectrum of affinities. In support of this latter hypothesis, BiP and hsc70 display marked differences in affinity for synthetic peptides as measured in a peptide-dependent ATPase reaction¹⁴⁸. In a small set of randomly chosen peptides, a range of at least 1,000-fold of Michaelis constant (K_m) values was obtained. Unfortunately, no pattern could be discerned that correlates any sequence or structural features of the peptides with their binding affinities.

In some cases (notably that of BiP and nascent secretory proteins), stress-70 proteins discriminate between folded and unfolded polypeptides, showing no propensity to associate with native protein structures. In other cases (for example DnaK with DNA replication complexes; hsc70 with clathrin cages) these proteins interact with apparently fully folded substrates and function to alter protein-protein contacts in multisubunit complexes. These schemes are not necessarily mutually exclusive, as DnaK and hsc70 also bind to unfolded polypeptides^{72,114}, and BiP may interact with protein components of the ER translocation system¹³⁸. To unify these observations, Rothman¹⁶ has suggested that the stress-70 proteins (and the other chaperones) be regarded as polypeptide-chain binding proteins (PCBs) with peptide-binding sites that can interact with chain segments only when they are part of incompletely folded structures or when they extend as loops from otherwise fully folded proteins. This hypothesis can explain how stress-70 proteins can interact with broad specificity with unfolded polypeptides whereas individual 'folded' proteins (such as clathrin light chains⁹²) can use a bait of extended chain to suborn the chaperone activity of a particular stress-70 family member to a specific purpose.

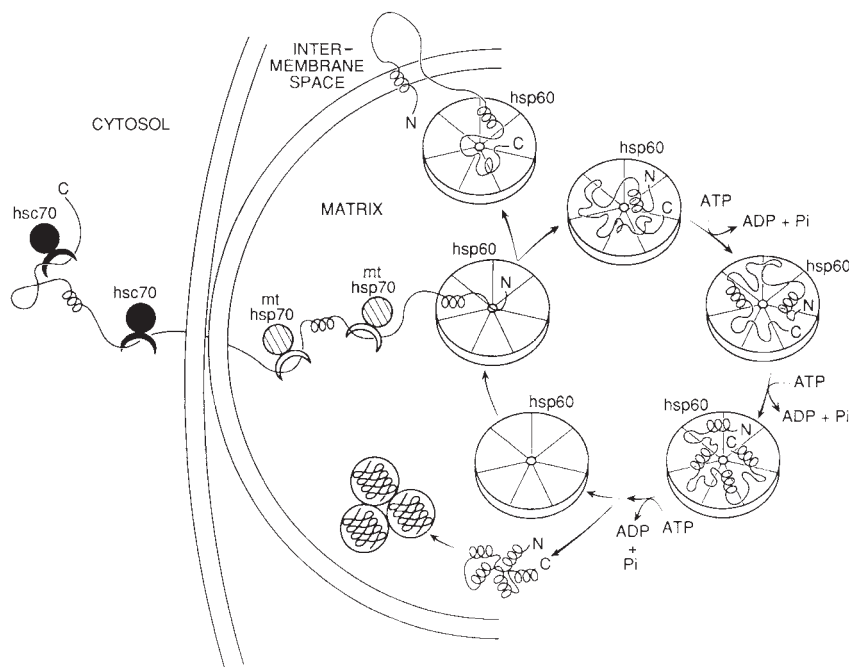


FIG. 2 Illustration of the proposed roles of hsp70 and chaperonin molecules during the import of mitochondrial precursors, their subsequent folding in the mitochondrial matrix and their reexport to the intermembrane space. The diagram is based on Fig. 1 from Neupert *et al.* (ref. 172) and in addition reflects multiple steps of folding on the chaperonin surface as suggested by Martin *et al.* (ref. 199).

Structural conservation of stress-70 proteins. Members of the stress-70 protein family have been highly conserved throughout evolution. DnaK, the single stress-70-related protein of *E. coli*, has about 50% amino-acid sequence identity with stress-70 proteins of eukaryotes¹⁵⁴, which are encoded by multiple *HSP70* genes that share between 50 and 95% identity at the nucleotide level⁶⁸. Comparison of all the known amino-acid sequences of stress-70 family members reveals that the N-terminal two-thirds (about 450 amino acids) of these proteins are much more highly conserved than the C-terminal portions, suggesting a conserved domain followed by a variable region^{155,156} (Fig. 3a). In addition, some stress-70 proteins contain short N-terminal or C-terminal extensions required for targeting to, or retention in, the appropriate cellular compartment. Thus BiP contains a cleavable N-terminal hydrophobic signal sequence specifying import into the ER^{79,125} and a C-terminal tetrapeptide (for example KDEL in mammalian BiPs, HDEL in yeast BiP) that is partly responsible for retention of the protein within the ER lumen¹⁵⁷. Similarly, stress-70 members located in mitochondria contain a hydrophilic N-terminal extension required for import into that organelle¹⁴¹.

ATP binding and hydrolytic activity are retained by a roughly 44K, N-terminal proteolytic fragment of bovine clathrin-uncoating ATPase (hsc70), although the ATPase activity is uncoupled from its normal dependence on clathrin binding¹⁵⁵. Similar N-terminal fragments are generated after proteolytic digestion of mammalian¹⁴⁹ and yeast⁷⁹ BiP proteins. The three-dimensional structure of the N-terminal fragment of bovine hsc70 has recently been solved to a resolution of 2.2 Å¹⁵⁸, revealing that the ATPase domain consists of two lobes with the nucleotide bound at the base of a deep cleft between them (Fig. 3b). Surprisingly, the folding topology of the hsc70 ATPase domain is nearly identical to that of the globular G actin monomer, despite there being little sequence homology between the two proteins¹⁵⁹. G actin monomers contain one molecule of noncovalently bound ATP, which is hydrolysed to form bound ADP and inorganic phosphate when G actin polymerizes to

form filamentous actin. The tertiary structure of the nucleotide-binding core of the hsc70 fragment is also similar to that of hexokinase, although the remainder of the structures of the two proteins are completely dissimilar¹⁵⁸.

The C-terminal domain of hsc70, which has been proposed¹⁵⁵ to be the 'specificity' domain that couples binding of target proteins to the ATPase activity of the conserved N-terminal domain, may 'dock' onto a face of the N-terminal domain that is lined with amino-acid residues that are highly conserved between stress-70 proteins¹⁵⁸. Although the structure of this portion of the molecule has not yet been determined for any stress-70 protein, two groups have now presented hypothetical models for the hsc70 C-terminal domain^{160,161}. Rippman *et al.*¹⁶⁰ deduced a consensus secondary structure for the C-terminal domains of 33 stress-70 proteins and obtained a pattern of helices and β -strands that could be aligned with that of the α -1 and α -2 domains of the human MHC class I antigen HLA. Flajnik *et al.* found that the same domains of an MHC class I protein from *Xenopus laevis*¹⁶² have a low amount of sequence identity with the C-terminal sequences of hsc70 and BiP proteins; they then showed that secondary structure predictions and hydrophathy analyses for the corresponding regions yield very similar results if a few gaps or insertions are introduced to optimize the alignment¹⁶¹. These findings prompted both groups to model the hsc70 C-terminal domain using the known three-dimensional structure of the human class I molecule¹⁶³ (Fig. 3c). The putative peptide binding cleft in each of the hypothetical structures is lined with both hydrophobic and hydrophilic residues. If the peptide binding domain of stress-70 proteins does indeed closely resemble that of HLA, it is very likely that the polypeptide chain would bind in an extended conformation¹⁶⁴. In fact, preliminary studies using NMR (nuclear magnetic resonance) indicate that DnaK binds a 13-residue synthetic peptide in a conformation that lacks any defined structural features¹⁶⁵.

A common mechanism for stress-70 action? From this abundance of disparate observations on the interaction of stress-70 proteins with their targets emerges a working model for a common mechanism for stress-70 action. Stress-70 proteins interact (probably through their C-terminal domains) with unfolded segments of polypeptide chain (Fig. 1). These unfolded segments may be presented either as nascent polypeptides emerging from the ribosome or from the lipid bilayer after membrane translocation, as sequences exposed by partial protein denaturation following an environmental stress, or as peptide loops extended from an otherwise native protein molecule. Sequence variability in the C-terminal domains of different family members may determine differences in the specificity of stress-70-peptide interactions. Although the basis for this specificity is not understood in any detail, it is likely that each stress-70 protein binds different polypeptide segments with a wide spectrum of affinities. Low-affinity binding may be reversed quickly and spontaneously, whereas release of peptide segments that are bound with high affinity may involve ATP hydrolysis mediated by the N-terminal domain of the stress-70 protein, or be effected through intervention of another cellular component (for example DnaJ, the NEM-sensitive factor, or both). Once released, the polypeptide chain has the opportunity to complete its folding by forming intramolecular interactions, or to assemble into oligomeric structures with nearby polypeptide chains, or to engage with the appropriate membrane translocation machinery or with another chaperone such as hsp60 (see below). If such interactions are not formed rapidly, stress-70 proteins, which are present at high concentration, may rebind and again stabilize the unfolded protein. During each interval of release, the polypeptide may either fold productively or form nonproductive intra- or intermolecular interactions yielding misfolded or aggregated molecules. It seems that the amino-acid sequences of wild-type proteins have evolved such that, under normal physiological conditions, productive folding or rebinding to stress-70 molecules are more likely events than misfolding. Thus wild-type

polypeptides can be maintained in an assembly competent state for very long periods in the absence of their appropriate homologous or heterologous partner subunits. Similarly, cycles of binding, release and rebinding will extend the period of interaction of stress-70 proteins with polypeptides that are unable to fold productively. Sequence alterations (amino-acid substitutions, deletions or insertions), aberrant post-translational modification (for example glycosylation) or conditions of stress (high temperature, altered redox potential, decreased Ca^{2+} concentration) would perturb the normal folding pathway and increase the probability of misfolding. In such circumstances intervention by stress-70 proteins may be able to delay, but not prevent, the formation of misfolded and/or aggregated structures that are dead-ends off the folding pathway. This aberrant folding could result in occlusion of the available sites on the polypeptide for the stress-70 protein, rendering the chaperone's action less effective. Such substoichiometric binding of BiP to aggregates of nonglycosylated haemagglutinin molecules has been observed in *in vivo* experiments¹³⁵.

It is obvious that any increased probability of misfolding could be reversed by increasing the local concentration of stress-70 protein, and it is equally clear that the cell has evolved mechanisms to sense increased amounts of nascent or unfolded proteins in different cellular compartments and to respond by inducing the transcription of the appropriate stress-70 gene. For example, in *E. coli* accumulation of unfolded proteins causes increased synthesis of DnaK (and other heat-shock proteins)^{113,166}. In eukaryotic cells, accumulation of unfolded proteins⁷⁶ or secretory precursors⁷⁹ in the cytosol results in induction of hsp70 and/or hsc70 proteins, whereas accumulation of unfolded proteins in the ER causes induction of BiP^{78,79}. Increased concentrations of individual, constitutively expressed stress-70 proteins can be achieved by accelerating their rates of synthesis; in cases where closely related family members can perform the same or similar tasks¹⁶⁷, synthesis of the constitutively expressed proteins may be augmented by *de novo* induction of closely related family members. In no case do we yet understand the nature of the induction signal generated by the presence of unfolded proteins or the pathways of their transduction to the nucleus (across the ER membrane in the case of BiP).

Finally, it has been suggested that stress-70 proteins assist correct polypeptide folding not only through their 'anti-folding' function but also by disentangling misfolded or aggregated proteins using the energy released during ATP hydrolysis^{87,120,168}. In cell-free translocation studies, precursors that have acquired defined structures after *in vitro* translation require unfolding for import into mitochondria^{2,169-171} or ER microsomes^{84,85,121}. As translocation in these systems is dependent on the presence of stress-70 proteins^{84-86,121} and ATP^{121,170}, it was suggested that the stress-70 proteins might be supplying the unfolding activity. But it is now thought¹⁷² that ATP hydrolysis may be required to release bound cytosolic stress-70 proteins and that the unfolding activity observed in cell-free extracts may be supplied by other components, such as the NEM-sensitive factor^{86,121} or proteins present on the mitochondrial surface¹⁷². The only stress-70 protein for which unfolding activity has been directly demonstrated *in vitro* is DnaK which can dissolve aggregates of heat-inactivated RNA polymerase⁹³. This process, which is dependent on the presence of hydrolysable ATP, requires at least stoichiometric amounts of DnaK. It is therefore possible that rather than actively unwinding the polypeptide chains in a catalytic process, DnaK might bind to peptide loops that are transiently exposed during 'breathing' of the structures of the misfolded proteins and, in a mechanism that parallels its putative role during normal folding, stabilize the polypeptide chain in a state competent for subsequent refolding to the correct conformation. ATP hydrolysis would then be required to promote release of DnaK to allow the polypeptide to continue the folding process.

The stress-90 protein family

Stress-90 proteins constitute a second major family of stress proteins (Table 1) whose members are present in all prokaryotic and eukaryotic organisms so far tested (reviewed in ref. 68). Members of the family show sequence conservation similar to that of the stress-70 family, there being greater than 40% identity between the various eukaryotic stress-90 proteins and the *E. coli* homologue Hsp90. These proteins are present in high abundance under normal growth conditions, but can be further induced by heat shock or other forms of stress. They are less numerous than the stress-70 proteins, there being only one member of the family in *E. coli* and *D. melanogaster*, and two members in *S. cerevisiae* that differ in their constitutive level and degree of inducibility⁶⁸. Vertebrate cells contain an additional stress-90 protein, variously named grp94 (ref. 173), ERp99 (ref. 174) or endoplasmic¹⁷⁵, which like BiP in the stress-70 family is synthesized as a precursor that contains an N-terminal signal sequence for ER translocation and a C-terminal KDEL tetrapeptide.

The cytosolic stress-90 proteins, whose apparent M_s vary from 87–92K, associate with a diverse range of cellular proteins including retroviral transforming proteins, steroid hormone receptors, cellular protein kinases, actin and tubulin (reviewed in ref. 68). The common feature of these interactions seems to be the stabilization of the target proteins in an inactive or unassembled state. Thus the association of monomeric hsp90 (and an unidentified 50K phosphoprotein) with pp60^{src} immedi-

ately after its synthesis stabilizes the transforming protein in an inactive state until it reaches its appropriate destination at the plasma membrane^{176,177}. Concomitant with its release from association with hsp90, pp60^{src} is phosphorylated on tyrosine, inserted into the plasma membrane, and activated as a kinase. Hsp90 is involved both in the initial folding of steroid hormone receptors¹⁷⁸ and in subsequent modulation of their DNA binding and transcriptional regulatory activities¹⁷⁹. Thus, aporeceptors synthesized in the absence of hsp90 lack transcriptional enhancement activity and responsiveness to hormonal activation. Under normal conditions, newly synthesized aporeceptors form a complex with a dimer of hsp90 in which the receptor is stabilized in a partially unfolded conformation that is unable to bind DNA (ref. 180). Binding of steroid hormone promotes dissociation of hsp90 from the complex and allows the receptor to bind DNA. When hsp90 is removed from the complex using high salt or temperature, the interaction of the receptor with DNA is unregulated, occurring in the presence or absence of hormone. Although binding of hsp90 has been mapped to a stretch of about 80 residues in the hormone binding site on the steroid receptor^{180–182}, the features that determine the specificity of binding and release of target polypeptides by members of the stress-90 family are not understood. The hsp90 binds ATP in a cation-dependent manner and undergoes autophosphorylation¹⁸³. But by contrast to stress-70 proteins and the chaperonins (see below), stress-90 proteins do not seem to catalyse the hydrolysis of ATP.

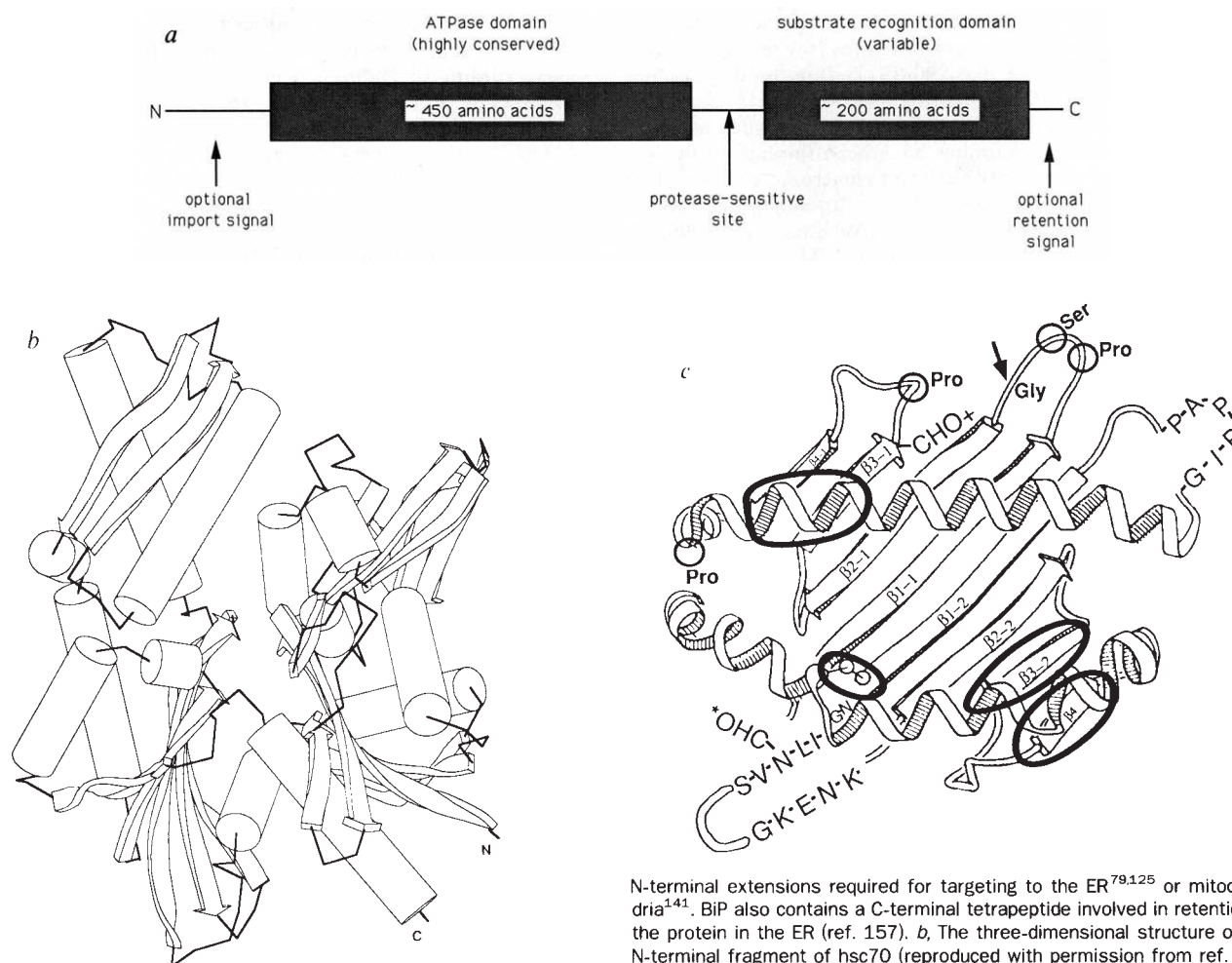


FIG. 3 Structure of stress-70 proteins. *a*, A linear diagram of a generalized stress-70 protein indicating two major domains: a highly conserved N-terminal domain that has ATPase activity and a less conserved C-terminal peptide-binding domain¹⁵⁵. Some stress-70 proteins also contain short

N-terminal extensions required for targeting to the ER^{79,125} or mitochondria¹⁴¹. BiP also contains a C-terminal tetrapeptide involved in retention of the protein in the ER (ref. 157). *b*, The three-dimensional structure of the N-terminal fragment of hsc70 (reproduced with permission from ref. 158) reveals that the ATPase domain consists of two structural lobes with the nucleotide bound at the base of a deep cleft between them. *c*, Possible structure of the C-terminal peptide binding domain of hsc70 (reproduced with permission from ref. 161) based on the known three-dimensional structure of the human MHC class I antigen HLA (ref. 163).

Nothing is known about the function of grp94, one of the most abundant proteins resident in the ER lumen. Like BiP, grp94 is induced by the accumulation of unfolded proteins in the ER (ref. 78), suggesting that it may function with BiP to assist the assembly of nascent polypeptides. Several abundant ER-resident proteins, including grp94 and BiP, are high capacity, low-affinity Ca^{2+} -binding proteins¹⁸⁴, but the functional significance of this property is not understood.

The GroEL/chaperonin family

The term chaperonin was suggested by Ellis⁸¹ to describe a class of molecular chaperones that are homologous in structure to *E. coli* GroEL. Members of this protein family are present in all prokaryotes and in those organelles of eukaryotic cells, such as mitochondria and chloroplasts, that have a probable endosymbiotic origin (Table 1). These proteins, which have been renamed chaperonin-60 (ref. 82), are large oligomers composed of 14 subunits each about 60K, arrayed as two stacked rings of seven subunits^{185,186}. In *E. coli*, GroEL interacts functionally in an ATP-dependent manner with GroES (chaperonin-10), a roughly 10K polypeptide that forms a single ring of seven subunits and is the second product of the GroE operon (reviewed in ref. 96). In the absence of unfolded protein substrates, the inherent ATPase activity of GroEL is inhibited by GroES^{187,188}. Mitochondria of mammalian cells contain a polypeptide that is structurally and functionally homologous to GroES (ref. 189, and P. Vütanen, unpublished results).

Remarkably, the general features of the interactions of chaperonin-60 molecules with their target polypeptides are very similar to those of stress-70 proteins, despite great differences in their sequences and oligomeric structures (for recent reviews and references see 67, 72, 96, 190–192). Thus both types of chaperones are highly abundant proteins whose rate of synthesis can be further induced by environmental stresses such as heat shock. Members of both families have been implicated in the assembly of nascent protein subunits into macromolecular structures, as well as in a number of other fundamental cellular processes. Chaperonin-60 molecules, like stress-70 proteins, bind ATP with high affinity and have weak ATPase activity and both types of proteins in some circumstances function together with other heat-shock proteins or cellular factors. Most importantly, both seem to act on their targets by stabilizing the conformation of folding intermediates, thereby preventing the formation of aberrant structures and directing the polypeptides down biologically productive assembly pathways.

GroEL. In *E. coli*, GroEL and GroES are abundant heat-shock proteins that are also required for viability under normal growth conditions⁷³. *GroE* mutants have phenotypes reminiscent of those of DnaK mutants⁷². Thus mutants lacking either GroEL or GroES have reduced rates of DNA and RNA synthesis, are blocked in cell division at nonpermissive temperatures, and show a reduction in overall protease activity. The GroE proteins are also required for bacteriophage morphogenesis in *E. coli*. Overproduction of both GroEL and GroES, but not of either alone, can suppress temperature-sensitive mutations in a large number of different genes¹⁹³ apparently by promoting the correct folding or assembly of the mutant polypeptides. Like stress-70 proteins, the GroE proteins also have a role in secretion. In bacterial cell-free protein translocation reactions, GroEL binds newly synthesized secretory precursors stabilizing them for membrane transit²⁴. *In vitro*, GroEL forms complexes with unfolded precursors of several secretory proteins including β -lactamase, proOmpA and prePhoE^{24,26}. *In vivo*, GroEL is required for export of β -lactamase but not other secretory precursors¹⁹⁴, perhaps because other prokaryotic chaperones such as trigger factor and secB function in its place²⁶. Finally, overproduction of GroEL in *E. coli* can facilitate the export of lacZ hybrid proteins¹¹⁵.

Insight into the mechanism of action of the *E. coli* GroE proteins has come from studies of their role in promoting

the folding and/or assembly of a number of enzymes, including prokaryotic ribulose biphosphate carboxylase (Rubisco)^{82,188,195,196}, pre- β -lactamase¹⁹⁷, citrate synthase¹⁹⁸, dihydrofolate reductase^{196,199} and rhodanese^{199,200}, and of GroEL itself²⁰¹. Partially folded protomers of these proteins form stable binary complexes with GroEL, in a process that competes both with biologically unproductive aggregation of the polypeptide chains and with their spontaneous refolding. GroEL does not interact with native proteins or with irreversibly denatured and aggregated molecules, but rather binds to labile folding intermediates likely to correspond to 'molten globules'^{14,15} or 'compact intermediates'¹³. In the absence of GroES, hydrolysis of ATP by GroEL promotes the discharge of the binary complex to release partially folded but catalytically inactive polypeptides. Whether release results in the generation of native, enzymatically active molecules depends on the nature of the polypeptide substrate and is related to the propensity of each polypeptide chain to fold spontaneously under the reaction conditions employed. If GroEL is available for rebinding after release, aggregation or continued folding of the polypeptides will be inhibited or delayed¹⁹⁹. In the presence of GroES, ATP hydrolysis-dependent folding occurs at the surface of GroEL through intermediate conformations that are progressively more compact but still enzymatically inactive¹⁹⁹. Finally, the polypeptide is released from the complex in a form that is apparently committed to completion of folding to the native state. In every case studied the overall effect of the coordinated action of the two GroE proteins is to increase the efficiency of refolding compared with that of the spontaneous process. But how the presence of the chaperonins influences the rate of the folding reaction varies significantly from protein to protein. Thus the rate of folding of Rubisco is enhanced 10-fold relative to the spontaneous process¹⁸⁸, whereas the rates of folding of citrate synthase¹⁹⁸ and pre- β -lactamase¹⁹⁷ are unchanged, and those of DHFR and rhodanese¹⁹⁹ are decreased. These differences are likely to be related to how the specific interaction between the polypeptide chain and the chaperonin promote or interfere with rate-limiting intramolecular interactions that normally take place during the spontaneous folding process.

No more than one or two molecules of unfolded polypeptide are bound to each oligomeric assembly of 14 GroEL molecules^{26,197,199}. This stoichiometry might suggest that the GroEL protomers in each heptameric ring, or in the tetradecamer, interact to form a single binding site. Alternatively, steric hindrance could limit the access of more than one or two protein molecules to 14 identical sites located in or near the hole in the centre of the doughnut-shaped structure^{185,186} (Fig. 4). A polypeptide could then be bound at up to 14 sites, each of which is capable of interacting with one of the multiple recognition motifs that may be exposed on a partially folded polypeptide. This second possibility is compatible with the suggestion that, in the presence of GroES, the folding of DHFR and rhodanese on the surface of GroEL occurs by progressive, ATP hydrolysis-dependent release of different portions of the bound polypeptide¹⁹⁹. The role of GroES would then be to modulate or coordinate the ATPase activity of each GroEL protomer, perhaps to prevent all sites discharging simultaneously leading to premature release of an only partially folded molecule¹⁹⁹. Overall the result of the interaction between GroEL and GroES would be to prolong contact between the chaperonin complex and its substrate as long as the polypeptide exposes structures recognized by GroEL.

Plastid Rubisco subunit binding protein (RBP). The chaperonin-60 present in chloroplasts of higher plants is a nuclear-encoded protein first identified because of its involvement in the assembly of the hetero-oligomeric plant Rubisco^{81,191,192}. In contrast to GroEL and mitochondrial hsp60s, each of which contain only one type of 60K subunits, RBP consists of two distinct but related subunits, α (61K) and β (60K), probably arranged as two layers of seven monomers each⁸¹. The wheat Rubisco small

subunit expressed in *E. coli* associates with GroEL, providing evidence for functional homology between GroEL and wheat RBP, which have 46% identity in amino-acid sequence. As several polypeptides associate with RBP after import into isolated chloroplasts²⁵ it seems likely that further studies will demonstrate the involvement of RBP in the assembly of plastid macromolecules other than Rubisco.

Mitochondrial hsp60. Mitochondria from *Tetrahymena*, yeast, maize, *Xenopus* and human cells also contain proteins that are related structurally and immunologically to the GroEL protein^{185,202}. A nuclear gene (*HSP60* or *MIF4*) encodes the mitochondrial homologue in *S. cerevisiae* (hsp60)²⁰³, which displays 54% and 43% amino-acid identity, respectively, with GroEL and the α component of chloroplast RBP. Hsp60 is expressed constitutively and is localized in the mitochondrial matrix²⁰⁴. In cells expressing a mutant (*mif4*) form of hsp60, subunits of mitochondrial enzymes fail to assemble into the appropriate macromolecular complexes despite being translocated into the mitochondrial matrix and undergoing normal processing in that compartment²⁰⁴. This is the case not only for proteins normally destined for the matrix, such as the β -subunit of F₁-ATPase and hsp60 itself²⁰⁵, but also for proteins with complex presequences, such as cytochrome *b*₂ and the Rieske Fe/S protein, whose final destination is the intermembrane space. These latter proteins accumulate as incompletely processed import intermediates²⁰⁴. Hsp60 has been directly implicated in folding of proteins in the mitochondrial matrix²⁰⁶. When assembly of precursors imported into wild-type mitochondria is arrested by depletion of ATP, by NEM treatment, or at low temperature, the unfolded polypeptides are associated with hsp60 in a high-*M_r* complex. Addition of ATP allows at least partial refolding of the polypeptides but does not promote their release from the complex. An unidentified factor seems to be required for the release reaction. Fig. 2 illustrates our current view of the role played by stress-70 and chaperonin molecules during the translocation and folding of polypeptides imported into mitochondria.

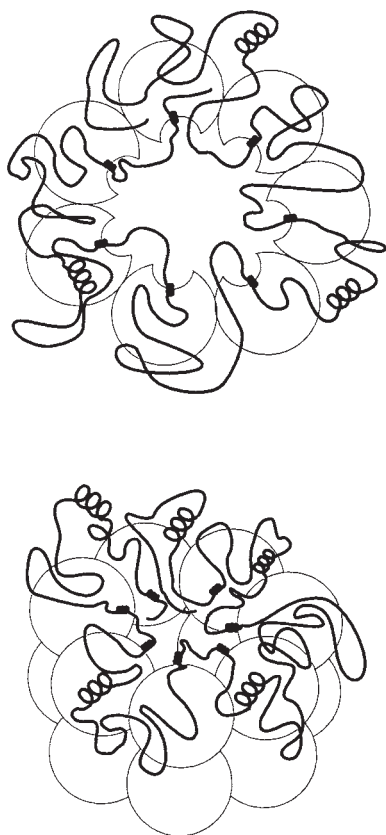


FIG. 4 Hypothetical model of the GroEL chaperonin structure reflecting multiple binding sites for a single polypeptide chain. The oligomeric structure is shown in views from the top and the side. The small corkscrews represent α -helices that may be recognition elements for binding to GroEL; the larger corkscrews represent general secondary structures in the bound polypeptide.

Mechanism of action of chaperonins? Although the broad features of the interaction of chaperonins with their polypeptide substrates have been illuminated by *in vitro* studies with GroEL and GroES (see above), many questions remain concerning the molecular details of the mechanism by which chaperonins promote protein folding and assembly. The quaternary structure of chaperonin oligomers has been revealed by electron microscopy^{185,186}, but nothing is known about the tertiary structures of chaperonin-60 or chaperonin-10 protomers or how they interact in the complex. Nor do we understand how chaperonins recognize their target polypeptides. The available evidence suggests that chaperonin-60 molecules bind to structural elements that are displayed by compact folding intermediates of a broad range of polypeptide substrates but absent or inaccessible in the native or aggregated forms of these proteins. The nature of these structural elements is not yet known, although preliminary experiments using NMR indicate that GroEL binds synthetic peptides that have the potential to form amphipathic α helices²⁰⁷. It will be important to learn whether chaperonin-60 molecules, like stress-70 proteins, display marked sequence specificity for binding. If so, recognition motifs on a folding protein may have evolved with a hierarchy of affinities that directs their order of release from the chaperonin complex. If so, rather than facilitating folding in a merely permissive fashion, the interaction between chaperonin and substrate may influence the pathway and the kinetics of the folding process. Finally, we need to understand the role that ATP hydrolysis plays in the folding and/or release of bound polypeptides and the manner in which chaperonin-10 regulates both this ATPase activity and the release reaction.

Roles of chaperonins and stress-70 proteins

Although stress-70 proteins and chaperonins share many common features in their modes of action, they do not perform interchangeable roles. Despite their coexistence in bacteria and in mitochondria and plastids of eukaryotic cells, each type of chaperones is independently essential for cell viability.

In mitochondria, both types of chaperones interact with unfolded or prefolded molecules: hsp70 molecules associate with imported polypeptides even before their translocation is completed¹⁴⁷, whereas hsp60 molecules become involved at a later stage^{90,204}. It is therefore possible that stress-70 molecules interact with less folded structural elements (perhaps segments of extended polypeptide chain¹⁶), whereas chaperonin-60 molecules associate with structural features common only to folding intermediates¹⁹⁶. Interestingly, two-dimensional NMR studies reveal that a synthetic peptide is bound by DnaK in an extended form, and by GroEL in an α -helical conformation¹⁶⁵.

Third, the two types of chaperones seem to play different roles during the process of polypeptide folding. Thus stress-70 molecules are thought to stabilize unfolded forms of their substrates; folding is envisaged as occurring after release from the chaperone. On the other hand, at least partial folding of polypeptides may take place on the surface of chaperonin-60 oligomers^{82,199,206}. This distinction may be a consequence of the dramatic difference in the oligomeric state of the two types of molecules. Stress-70 proteins are thought to contain a single peptide-binding site and to associate as monomers with their target polypeptides⁹¹. Therefore, any folding step that is inhibited by chaperone binding must occur after dissociation of the complex. By contrast, the functional form of the chaperonin-60 molecule is an oligomer of 14 subunits, which apparently binds no more than 1 or 2 target polypeptides^{26,197,199}. Multivalent binding of a protein to the chaperonin-60 oligomer would allow release and partial folding of one portion of the chain while the polypeptide remains attached at other sites. The extent of folding of any protein on the chaperonin-60 surface might then depend on the role in the folding process of the portion of the chain that is bound with highest affinity.

Fourth, although stress-70 proteins such as DnaK and hsc70 are known to interact both with unfolded polypeptides and with protein oligomers (such as replication complexes or clathrin triskelions), evidence is lacking for any *in vivo* role for chaperonins in the rearrangement of oligomeric complexes.

Finally, although stress-70 and chaperonin-60 molecules function alongside each other in bacteria and in the matrix of mitochondria and plastids, chaperonins have not been identified in other compartments of eukaryotic cells. Whether stress-70 family members manage alone in supporting polypeptide folding and assembly in those compartments, or whether other proteins fulfil the function of the chaperonins remains to be determined.

Future directions

We now appreciate that chaperones are involved at all stages of cellular metabolism, during protein biosynthesis and maturation, in protection from environmental stress, in rearrangements of cellular macromolecules during functional cycles of assembly and disassembly, and finally in targeting proteins for degradation. Much progress has been made in characterizing chaperones that are members of three families of major stress proteins, and in identifying a number of unrelated proteins that also regulate or facilitate polypeptide folding in the cell. The major challenge now lies in elucidating the specific molecular mechanisms by which chaperones recognize their target proteins and promote, inhibit or reverse folding and assembly.

Although acceptance of the involvement of chaperones has required some revision of long-held views about the spontaneity of the process of protein folding, there was less resistance to a role during folding for enzymes such as PDI and the PPIases. Despite advances in our understanding of their mechanisms of action, important questions remain to be answered about the *in vivo* function and substrates of these enzymes and of a variety of newly discovered related proteins. □

Mary-Jane Gething and Joseph Sambrook are at the Department of Biochemistry, University of Texas Southwestern Medical Centre, 5323 Harry Hines Boulevard, Dallas, Texas 75235, USA. M.J.G. is an Investigator of the Howard Hughes Medical Institute.

39. Geetha-Habib, M., Noiva, R., Kaplan, H. A. & Lennarz, W. J. *Cell* **54**, 1053-1060 (1988).
40. Bulleid, N. J. & Freedman, R. B. *EMBO J.* **9**, 3527-3532 (1990).
41. Wettlaufer, J. R., Combs, K., Spinner, S. N. & Joiner, B. J. *J. Biol. Chem.* **265**, 9800-9807 (1990).
42. Lewis, M. J., Mazzarella, R. A. & Green, M. J. *J. Biol. Chem.* **260**, 3050-3057 (1985).
43. Mazzarella, R. A., Srinivasan, M., Haugejorden, S. M. & Green, M. J. *J. Biol. Chem.* **265**, 1094-1101 (1990).
44. Russel, M. *Molec. Microbiol.* **5**, 1607-1613 (1991).
45. Schreiber, S. L. *Science* **251**, 283-287 (1991).
46. Fesik, S. W. *et al. Science* **250**, 1406-1409 (1990).
47. Van Duyn, G. D., Standaert, R. F., Karplus, P. A., Schrieber, S. L. & Clardy, J. *Science* **252**, 839-842 (1991).
48. Liu, J. & Walsh, C. T. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4028-4032 (1990).
49. Hayano, T., Takahashi, N., Kato, S., Manki, N. & Suzuki, M. *Biochemistry* **30**, 3041-3048 (1991).
50. Tropschug, M. *et al. J. Biol. Chem.* **263**, 14433-14440 (1988).
51. Koser, P. L., Sylvester, D., Livi, G. P. & Bergsma, D. J. *Nucleic Acids Res.* **18**, 1643 (1990).
52. Shieh, B. H., Stammes, M. A., Seavello, S., Harris, G. L. & Zuker, C. S. *Nature* **338**, 67-70 (1989).
53. Schneuwly, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 5390-5394 (1989).
54. Price, E. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 1903-1907 (1991).
55. Spik, G. *et al. J. Biol. Chem.* **266**, 10735-10738 (1991).
56. Caroni, P., Rothenfluh, A., McGlynn, E. & Schneider, C. J. *J. Biol. Chem.* **266**, 10739-10742 (1991).
57. Hasel, K. W., Glass, J. R., Goudout, M. & Sutcliffe, J. G. *Molec. cell. Biol.* **11**, 3484-3491 (1991).
58. Jin, Y.-J. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 6677-6681 (1991).
59. Haendler, D., Hofer-Warbinek, R. & Hofer, E. *EMBO J.* **6**, 947-950 (1987).
60. Danielson, P. E. *et al. DNA* **7**, 261-267 (1988).
61. Tropschug, M., Barthelmess, I. B. & Neupert, W. *Nature* **342**, 953-955 (1989).
62. Wilderrecht, G., Brizuela, L., Elliston, K., Sigal, N. H. & Siekierka, J. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1029-1033 (1991).
63. Heitman, J., Movva, N. R., Hiestand, P. C. & Hall, M. N. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1948-1952 (1991).
64. Stammes, M. A., Shieh, B.-H., Chuman, L., Harris, G. L. & Zuker, C. S. *Cell* **65**, 219-227 (1991).
65. Harrison, R. K. & Stein, R. L. *Biochemistry* **29**, 3813-3816 (1990).
66. Michnick, S. W., Rosen, M. K., Wandless, T. J., Karplus, M. & Schreiber, S. L. *Science* **252**, 836-839 (1991).
67. Ellis, R. J. & Hemmingsen, S. M. *Trends biochem. Sci.* **14**, 339-342 (1989).
68. Craig, E. A. *Rev. Genet.* **22**, 631-677 (1988).
69. Craig, E. A. *CRC Crit. Rev. Biochem.* **18**, 239-280 (1985).
70. Lindquist, S. A. *rev. Biochem.* **55**, 1151-1191 (1986).
71. Craig, E. A. & Gross, C. A. *Trends biochem. Sci.* **16**, 135-140 (1991).
72. Georgopoulos, C., Ang, D., Liberek, K. & Zyllic, M. in *Stress Proteins in Biology and Medicine* (eds Morimoto, R. I., Tissières, A. & Georgopoulos, C.) 191-221 (Cold Spring Harbor, New York, 1990).
73. Fayet, O., Zieglerhoffer, T. & Georgopoulos, C. *J. Bact.* **171**, 1379-1385 (1989).
74. Subjeck, J. R. & Shyy, T.-T. *Am. J. Physiol.* **250**, C1-C17 (1986).
75. Lee, A. S. *Trends biochem. Sci.* **12**, 20-23 (1987).
76. Ananthan, J., Goldberg, A. L. & Voellmy, R. *Science* **232**, 522-524 (1986).
77. Hiromi, Y. & Hotta, Y. *EMBO J.* **4**, 1681-1687 (1985).
78. Kozutsumi, Y., Segal, M., Normington, K., Gething, M. J. & Sambrook, J. *Nature* **332**, 462-464 (1988).
79. Normington, K., Kohno, K., Kozutsumi, Y., Gething, M. J. & Sambrook, J. *Cell* **57**, 1223-1236 (1989).
80. Werner-Washburne, M., Stone, D. E. & Craig, E. A. *Molec. cell. Biol.* **7**, 2568-2577 (1987).
81. Hemmingsen, S. M. *et al. Nature* **333**, 330-334 (1988).
82. Golubovoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. *Nature* **342**, 884-889 (1989).
83. Beckmann, R. P., Mizzen, L. A. & Welch, W. J. *Science* **248**, 850-854 (1990).
84. Chirico, W. J., Waters, M. G. & Blobel, G. *Nature* **332**, 805-809 (1988).
85. Deshaies, R. J., Koch, B. D., Werner-Washburne, M., Craig, E. A. & Schekman, R. *Nature* **332**, 800-805 (1988).
86. Murakami, H., Pain, D. & Blobel, G. *J. Cell Biol.* **107**, 2051-2057 (1988).
87. Pelham, H. R. B. *Cell* **46**, 959-961 (1986).
88. Hendershot, L. M., Bole, D. & Kearney, J. F. *Immun. Today* **8**, 111-114 (1987).
89. Ng, D. T. W., Randall, R. E. & Lamb, R. A. *J. Cell Biol.* **109**, 3273-3289 (1989).
90. Kang, P. J. *et al. Nature* **348**, 137-143 (1990).
91. Rothman, J. E. & Schmid, S. L. *Cell* **46**, 5-9 (1986).
92. DeLuca-Flaherty, C., McKay, D. B., Parham, P. & Hill, B. L. *Cell* **62**, 875-887 (1990).
93. Skowry, D., Georgopoulos, C. & Zyllic, M. *Cell* **62**, 939-944 (1990).
94. Chiang, H.-L., Terleky, S. R., Plant, C. P. & Dice, J. F. *Science* **246**, 382-385 (1989).
95. Dice, J. F. *Trends biochem. Sci.* **376**, 305-309 (1990).
96. Georgopoulos, C. & Ang, D. *Semin. Cell. Biol.* **1**, 19-25 (1990).
97. Hartl, F. U. & Neupert, W. *Science* **247**, 930-938 (1990).
98. Hallberg, R. L. *Semin. Cell Biol.* **1**, 37-45 (1990).
99. Hemmingsen, S. M. *Semin. Cell Biol.* **1**, 47-54 (1990).
100. Laskey, R. A., Honda, B. M. & Finch, J. T. *Nature* **275**, 416-420 (1978).
101. Dingwall, C. & Laskey, R. A. *Semin. Cell Biol.* **1**, 11-17 (1990).
102. Bonifacio, J. S. *et al. J. Biol. Chem.* **263**, 8965-8971 (1988).
103. Pettet, C. L., Alarcon, B., Malin, R., Weinberg, K. & Terhorst, C. *J. Biol. Chem.* **262**, 4854-4859 (1987).
104. Degen, E. & Williams, D. B. *J. Cell Biol.* **112**, 1099-1115 (1991).
105. Collier, D. N., Bankaitis, V. A., Weiss, J. B. & Bassford, P. J. *Jr Cell* **53**, 273-283 (1988).
106. Crooke, E., Guthrie, B., Lecker, S., Lill, R. & Wickner, W. *Cell* **54**, 1003-1011 (1988).
107. Kumamoto, C. A. *Molec. Microbiol.* **5**, 19-22 (1991).
108. Holmgren, A. & Branden, C.-I. *Nature* **342**, 248-251 (1989).
109. Sambrook, J. F. & Gething, M. J. *Nature* **342**, 224-225 (1989).
110. Prevelige, P., Thomas, D. & King, J. *J. molec. biol.* **202**, 743-757 (1988).
111. Pelham, H. R. B. in *Stress Proteins in Biology and Medicine* (eds Morimoto, R. I., Tissières, A. & Georgopoulos, C.) 287-299 (Cold Spring Harbor, New York, 1990).
112. Morimoto, R. I., Tissières, A. & Georgopoulos, C. in *Stress Proteins in Biology and Medicine* 1-36 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1990).
113. Parsell, D. A. & Sauer, R. T. *Genes Dev.* **3**, 1226-1232 (1989).
114. Clarke, C. F. *et al. Molec. cell. Biol.* **3**, 1206-1215 (1988).
115. Phillips, G. J. & Silhavy, T. J. *Nature* **344**, 882-884 (1990).
116. Zyllic, M., Lebowitz, J. H., McMacken, R. & Georgopoulos, C. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6431-6435 (1983).
117. Liberek, K., Georgopoulos, C. & Zyllic, M. *Proc. natn. Acad. Sci. U.S.A.* **18**, 6632-6636 (1988).
118. Wickner, S., Hoskins, J. & McKenney, K. *Nature* **350**, 165-167 (1991).
119. Liberek, K., Marszałek, J., Ang, D., Georgopoulos, C. & Zyllic, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2874-2878 (1991).
120. Lewis, M. J. & Pelham, H. R. B. *EMBO J.* **4**, 3137-3143 (1985).
121. Zimmermann, R., Sagstetter, M., Lewis, M. J. & Pelham, H. R. B. *EMBO J.* **7**, 2875-2880 (1988).
122. Chappell, T. G. *et al. Cell* **45**, 3-13 (1986).
123. Schlossmann, D. M., Schmid, S. L., Braell, W. A. & Rothman, J. E. *J. Cell Biol.* **99**, 723-733 (1984).
124. Dice, J. F. & Chiang, H.-L. *Biochem. Soc. Symp.* **55**, 45-55 (1989).

1. Randall, L. L. & Hardy, S. J. S. *Cell* **46**, 921-928 (1986).
2. Eilers, M. & Schatz, G. *Nature* **322**, 228-232 (1986).
3. Gething, M. J., McCammon, K. & Sambrook, J. *Cell* **46**, 939-950 (1986).
4. Gething, M. J., McCammon, K. & Sambrook, J. *Meth. Cell Biol.* **32**, 185-206 (1989).
5. Bernstein, F. C. *et al. J. molec. Biol.* **112**, 535-542 (1977).
6. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzie, T. F. & Weng, J. in *Crystallographic Data Bases: Information Content, Software Systems, Scientific Applications* (eds Allen, F. H., Bergerhoff, G. & Sievers, R.) 107-132 (Data Commission of the International Union of Crystallography, Bonn, Cambridge and Chester, 1987).
7. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. *Proc. natn. Acad. Sci. U.S.A.* **47**, 1309-1314 (1961).
8. Anfinsen, C. B. *Science* **181**, 223-230 (1973).
9. Baldwin, R. L. *Trends biochem. Sci.* **14**, 291-294 (1989).
10. Fischer, G. & Schmid, F. X. *Biochemistry* **29**, 2205-2212 (1990).
11. Kim, P. S. & Baldwin, R. L. A. *Rev. Biochem.* **59**, 631-660 (1990).
12. Jaenicke, R. *Prog. Biophys. molec. Biol.* **49**, 117-237 (1987).
13. Creighton, T. E. *Biochem. J.* **270**, 1-16 (1990).
14. Kuwajima, K. *Proteins struct. funct. Genet.* **6**, 87-103 (1989).
15. Pitsyn, O. B. *J. Prot. Chem.* **6**, 272-293 (1991).
16. Rothman, J. E. *Cell* **59**, 591-601 (1989).
17. Freedman, R. B. *Trends biochem. Sci.* **9**, 438-441 (1984).
18. Colman, A., Besley, J. & Valle, G. J. *molec. Biol.* **160**, 459-474 (1982).
19. Hendershot, L. M. *J. Cell Biol.* **111**, 829-837 (1990).
20. Copeland, C. S., Doms, R. W., Bolza, E. M., Webster, R. G. & Helenius, A. *J. Cell Biol.* **103**, 1179-1191 (1986).
21. Mitraki, A. & King, J. *BioTechnology* **7**, 690-697 (1989).
22. Haas, I. G. & Wabl, M. *Nature* **306**, 387-389 (1983).
23. Bole, D. G., Hendershot, L. M. & Kearney, J. F. *J. Cell Biol.* **102**, 1558-1566 (1986).
24. Bochkareva, E. S., Lissin, N. M. & Girshovich, A. S. *Nature* **336**, 254-257 (1988).
25. Lubben, T. H., Donaldson, E. K., Viitanen, P. V. & Gatenby, A. A. *Pl. Cell* **1**, 1223-1230 (1989).
26. Lecker, S. *et al. EMBO J.* **8**, 2703-2709 (1989).
27. Freedman, R. B., Bulleid, N. J., Hawkins, H. C. & Paver, J. L. *Biochem. Soc. Symp.* **55**, 167-192 (1989).
28. Freedman, R. B. *Cell* **57**, 1069-1072 (1989).
29. Lang, K., Schmid, F. X. & Fischer, G. *Nature* **329**, 268-270 (1987).
30. Schönbrunner, E. R. *et al. J. Biol. Chem.* **266**, 3630-3635 (1991).
31. Bachinger, H. P. & Compton, L. A. *J. cell. Biochem. suppl.* **156**, 188 (1991).
32. Roth, R. A. & Pierce, S. B. *Biochemistry* **26**, 4179-4182 (1987).
33. Bulleid, N. J. & Freedman, R. B. *Nature* **335**, 649-651 (1988).
34. Lamantia, M. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 4453-4457 (1991).
35. Farquhar, R. *et al. Gene* (in press).
36. Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A. & Rutter, W. J. *Nature* **317**, 267-270 (1985).
37. Holmgren, A. A. *Rev. Biochem.* **54**, 237-271 (1985).
38. Freedman, R. B., Hawkins, H. C., Murrant, S. J. & Reid, L. *Biochem. Soc. Trans.* **16**, 96-99 (1988).

125. Munro, S. & Pelham, H. R. B. *Cell* **46**, 291–300 (1986).
126. Rose, M. D., Misra, L. M. & Vogel, J. P. *Cell* **57**, 1211–1221 (1989).
127. Fontes, E. B. P. et al. *Pl. Cell* (in the press).
128. Morrison, S. L. & Scharff, M. D. *J. Immun.* **114**, 655–659 (1975).
129. Pouyssegur, J., Shiu, R. P. C. & Pastan, I. *Cell* **11**, 941–947 (1977).
130. Conde, J. & Fink, G. R. *Proc. natn. Acad. Sci. USA* **73**, 3651–3655 (1976).
131. Polaina, J. & Conde, J. *Molec. Gen. Genet.* **186**, 253–258 (1982).
132. Boston, R. S., Fontes, E. B. P., Shark, B. B. & Wrobel, R. L. *Pl. Cell* (in the press).
133. Dörner, A. J., Bole, D. G. & Kaufman, R. J. *J. Cell Biol.* **105**, 2665–2674 (1987).
134. Blount, P. & Merlie, J. P. *J. Cell. Biol.* **113**, 1125–1132 (1991).
135. Hurtley, S. M., Bole, D. G., Hoover-Litty, H., Helenius, A. & Copeland, C. S. *J. Cell Biol.* **108**, 2117–2126 (1989).
136. Nicholson, R. C., Williams, D. B. & Moran, L. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1159–1163 (1990).
137. Hendershot, L. M., Bole, D., Köhler, G. & Kearney, J. F. *J. Cell Biol.* **104**, 761–767 (1987).
138. Vogel, J. P., Misra, L. M. & Rose, M. D. *J. Cell Biol.* **110**, 1885–1895 (1990).
139. Rothblatt, J. A., Deshaies, R. J., Sanders, S. L., Daum, G. & Schekman, R. *J. Cell Biol.* **109**, 2641–2652 (1989).
140. Sadler, I. et al. *J. Cell. Biol.* **109**, 2665–2675 (1989).
141. Craig, E. A. et al. *Molec. cell. Biol.* **9**, 3000–3008 (1989).
142. Amir-Shipra, D., Leustek, T., Dalie, B., Weissbach, H. & Brot, N. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1749–1752 (1990).
143. Engman, D. M., Kirchoff, L. V. & Doneison, J. E. *Molec. cell. Biol.* **9**, 5163–5168 (1989).
144. Leustek, T., Dalie, B., Amir-Shipra, D., Brot, N. & Weissbach, H. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7805–7808 (1989).
145. Mizzen, L. A., Chang, C., Garrels, J. I. & Welch, W. J. *J. biol. Chem.* **264**, 20664–20675 (1989).
146. Craig, E. A., Kramer, J. & Kosic-Smithsers, J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4156–4160 (1987).
147. Scherer, P. E., Krieg, U. C., Hwang, S. T., Vestweber, D. & Schatz, G. *EMBO J.* **9**, 4315–4322 (1990).
148. Flynn, G. C., Chappell, T. G. & Rothman, J. E. *Science* **245**, 385–390 (1989).
149. Kassenbrock, C. K. & Kelly, R. B. *EMBO J.* **8**, 1461–1467 (1989).
150. Schmid, S. L., Braell, W. A. & Rothman, J. E. *J. biol. Chem.* **260**, 10057–10062 (1985).
151. Sharma, S., Rodgers, L., Brandsma, J., Gething, M. J. & Sambrook, J. *EMBO J.* **4**, 1479–1489 (1985).
152. Bird, P., Gething, M. J. & Sambrook, J. *J. Cell. Biol.* **105**, 2905–2914 (1987).
153. Wilson, I. A., Skehel, J. J. & Wiley, D. C. *Nature* **289**, 366–373 (1981).
154. Bardwell, J. C. A. & Craig, E. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 848–852 (1984).
155. Chappell, T. G., Konforti, B. B., Schmid, S. L. & Rothman, J. E. *J. biol. Chem.* **262**, 746–751 (1987).
156. Milarski, K. L. & Morimoto, R. I. *J. Cell Biol.* **109**, 1947–1962 (1989).
157. Pelham, H. R. B. *A. Rev. Cell Biol.* **5**, 1–23 (1989).
158. Flaherty, K. M., DeLuca-Flaherty, C. & McKay, D. B. *Nature* **346**, 623–628 (1990).
159. Flaherty, K. M., McKay, D. B., Kabash, W. & Holmes, K. C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5041–5045 (1991).
160. Rippmann, F., Taylor, W. R., Rothbard, J. B. & Green, N. M. *EMBO J.* **10**, 1053–1059 (1991).
161. Flajnik, M. F., Canel, C., Kramer, J. & Kasahara, M. *Immunogenetics* **33**, 295–300 (1991).
162. Flajnik, M. F., Canel, C., Kramer, J. & Kasahara, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 537–541 (1991).
163. Bjorkman, P. J. et al. *Nature* **329**, 506–512 (1987).
164. Madden, D. R., Gorga, J. C., Strominger, J. F. & Wiley, D. C. *Nature* **353**, 321–325 (1991).
165. Landry, S. J. & Gierasch, L. M. *Peptides* (in the press).
166. Goff, S. A. & Goldberg, A. L. *Cell* **41**, 587–595 (1985).
167. Craig, E. A. in *Stress Proteins in Biology and Medicine* (eds Morimoto, R. I., Tissières, A. & Georgopoulos, C.) 320–321 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1990).
168. Rothman, J. E. & Kornberg, R. D. *Nature* **322**, 209–210 (1986).
169. Eilers, M., Hwang, S. & Schatz, G. *EMBO J.* **7**, 1139–1145 (1988).
170. Eilers, M. & Schatz, G. *Cell* **52**, 481–483 (1988).
171. Chen, W.-J. & Douglas, M. G. *J. biol. Chem.* **262**, 15605–15609 (1987).
172. Neupert, W., Hartl, F.-U., Craig, E. A. & Pfanner, N. *Cell* **63**, 447–450 (1990).
173. Lee, A. S., Bell, J. & Ting, J. *J. biol. Chem.* **259**, 4616–4621 (1984).
174. Mazzarella, R. A. & Green, M. J. *J. biol. Chem.* **262**, 8875–8883 (1987).
175. Koch, G. L. E., Smith, M., Mace, D. R. J., Webster, P. & Mortara, R. *J. Cell Sci.* **86**, 217–232 (1986).
176. Brugge, J. S., Erikson, E. & Erikson, R. *Cell* **25**, 363–372 (1981).
177. Brugge, J. S., Yonemoto, W. & Darrow, D. *Molec. cell. Biol.* **4**, 2697–2704 (1983).
178. Picard, D. et al. *Nature* **348**, 166–168 (1990).
179. Picard, D., Salsner, S. J. & Yamamoto, K. R. *Cell* **54**, 1073–1080 (1988).
180. Dalman, F. C. et al. *J. biol. Chem.* **264**, 19815–19821 (1989).
181. Howard, K. J., Holley, S. J., Yamamoto, K. R. & Distelhorst, C. W. *J. biol. Chem.* **265**, 11928–11935 (1990).
182. Dalman, F. C., Scherrer, L. C., Taylor, L. P., Akil, H. & Pratt, W. B. *J. biol. Chem.* **266**, 3482–3490 (1991).
183. Cserrmely, P. & Kahn, C. R. *J. biol. Chem.* **266**, 4943–4950 (1991).
184. Mace, D. R. J. & Koch, G. L. E. *J. Cell. Sci.* **91**, 61–70 (1988).
185. McMullin, T. W. & Hallberg, R. L. *Molec. cell. Biol.* **8**, 371–380 (1988).
186. Hendrix, R. W. *J. molec. Biol.* **129**, 375–392 (1979).
187. Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R. & Georgopoulos, C. *J. biol. Chem.* **261**, 12414–12419 (1986).
188. Vitanen, P. V. et al. *Biochemistry* **29**, 5665–5671 (1990).
189. Lubben, T. H., Gatenby, A. A., Donaldson, G. K., Lorimer, G. H. & Viitanen, P. V. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7683–7687 (1990).
190. Ellis, R. J., van der Vies, S. M. & Hemmingsen, S. M. *Biochem. Soc. Symp.* **55**, 145–153 (1989).
191. Ellis, R. J. *Science* **250**, 954–959 (1990).
192. Gatenby, A. A. & Ellis, R. J. *Rev. Cell Biol.* **6**, 125–149 (1990).
193. Van Dyk, T. K., Gatenby, A. A. & LaRossa, R. A. *Nature* **342**, 451–453 (1989).
194. Kusukawa, N., Yura, T., Chiharu, U., Akiyama, Y. & Ito, K. *EMBO J.* **8**, 3517–3521 (1987).
195. Goloubinoff, P., Gatenby, A. A. & Lorimer, G. H. *Nature* **337**, 44–47 (1989).
196. Lorimer, G. H., Viitanen, P. V., van der Vies, S. M., Donaldson, G. & Gatenby, A. A. *J. cell. Biochem. suppl.* **15G**, 160 (1991).
197. Laminet, A. A., Zieglerhoffer, T., Georgopoulos, C. & Plückthun, A. *EMBO J.* **9**, 2315–2319 (1990).
198. Buchner, J. et al. *Biochemistry* **30**, 1586–1591 (1991).
199. Martin, J. et al. *Nature* **352**, 36–42 (1991).
200. Mendoza, J. A., Rogers, E., Lorimer, G. H. & Horowitz, P. M. *J. biol. Chem.* **266**, 13044–13049 (1991).
201. Lissin, N. M., Vanyaminov, S. Y. & Girshovich, A. S. *Nature* **348**, 339–342 (1991).
202. McMullin, T. W. & Hallberg, R. L. *Molec. cell. Biol.* **7**, 4414–4423 (1987).
203. Reading, D. S., Hallberg, R. L. & Myers, A. M. *Nature* **337**, 655–659 (1989).
204. Cheng, M. Y. et al. *Nature* **337**, 620–625 (1989).
205. Cheng, M. Y., Hartl, F.-U. & Horwich, A. L. *Nature* **348**, 455–458 (1990).
206. Ostermann, J., Horwich, A. L., Neupert, W. & Hartl, F.-U. *Nature* **341**, 125–130 (1989).
207. Landry, S. J. & Gierasch, L. M. *Biochemistry* **30**, 7359–7362 (1991).

Ridges, hotspots and their interaction as observed in seismic velocity maps

Yu-Shen Zhang* & Toshiro Tanimoto

Seismological Laboratory, 252-21, California Institute of Technology, Pasadena, California 91125, USA

A new global S-wave velocity model reveals that although mid-ocean ridges and hotspots are both underlain by low-velocity anomalies in the mantle, these have distinctly different structures. This implies that there are differences between the upwelling mechanisms under ridges and under hotspots. The velocity model also shows that there may be interactions between ridges and hotspots near Afar and St Helena.

RIDGES and hotspots are two main forms of upwelling from the interior of the Earth and are essential features of global tectonics. In ocean basins, they are the dominant modes of igneous activity, and are known to produce magmas of differing, although sometimes overlapping, geochemistry¹. The mechanisms operating below these features have been inferred from surface observations, including geochemical, topographic and

geoid data, and theoretical models based on gross structure and evolution of the mantle^{2–5}. Seismic data provide critical information for understanding these features from maps of three-dimensional (3D) structure. On a local scale, seismic techniques have been successful, for example, in constraining the size of magma chambers under ridges⁶, but such studies are restricted to particular areas where dense observations have been made.

We have obtained a new global, 3D S-wave velocity model for the upper mantle by analysis of travel times of long-period Love and Rayleigh waves (75–250 s) using ~18,000 seismograms. It is now becoming possible for global seismic studies to detect seismic velocity anomalies caused by various tectonic features (as opposed to purely long-wavelength variations), and to make comparisons between regions in order to understand the underlying mechanisms. Details of the data analysis and S-wave velocity structure are reported elsewhere^{7,8}. Here we focus on S-wave velocity variation under ridges and hotspots.

Long-wavelength features of the new model are similar to previous global, 3D seismic models^{9–14}. But previous models were represented by spherical harmonic expansion only up to angular degrees 6–10, with horizontal resolution lengths of 4,000–6,000 km. The chief advance of our model, which is expanded up to degree 36, is a considerable improvement in the

* Present address: Institute of Tectonics, University of California, Santa Cruz, California 95064, USA.

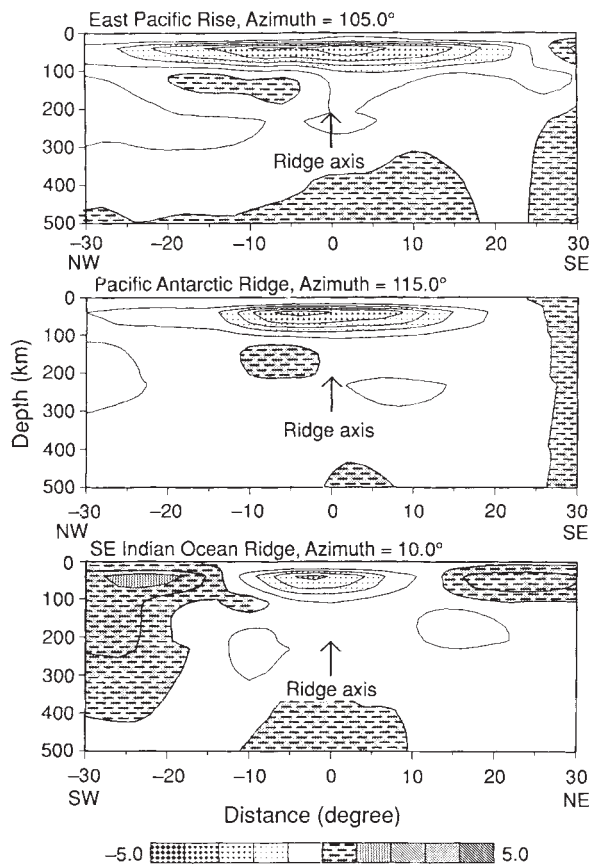


FIG. 1 Depth slices of S-wave velocity anomaly for three ridges. From top to bottom: East Pacific Rise, Pacific–Antarctic Ridge and Southeast Indian Ocean Ridge. Values are averaged over 30° along the strike of ridge axes. The ridge axes are aligned along the centre. Three hatched areas in Fig. 2 (except that in the South Atlantic) show the locations of these areas. Velocity anomalies are given in per cent of the global average at each depth and indicated by the shading. Contours are given every 0.5%. Abscissa: distance from ridge axis in degrees ($1^\circ = 111$ km). Errors associated with velocity variations are 0.4–0.6%, so each contour interval represents roughly 1 s.d. Note that, regardless of spreading rates, very slow velocity anomalies ($< -1\%$) are confined to shallow depths. On the other hand, -1% contours are elongated in the direction of spreading and correlated with spreading rates.

resolution. Although the resolution varies slightly from place to place, the mean lateral averaging length is ~ 800 – $1,000$ km. The depth averaging length is ~ 70 – 80 km near the surface and 300 km at a depth of 500 km. This resolution allows us to probe into relatively fine details of ridges and hotspots.

Because S-wave velocities are sensitive to temperature variations and degree of partial melting, S-wave velocity results provide critical constraints on models of amount and depth of partial melting, temperature anomalies and discharge rates at ridges and hotspots. In our upper-mantle S-wave velocity model, we find that mid-ocean ridges and hotspots are associated with regions of low seismic velocity, but the depth extent of the low-velocity anomalies is distinctly different. This is true not only in each type of region, but also in mixed regions where hotspots exist on or very near a mid-ocean ridge axis. A typical mixed region is the Mid-Atlantic Ridge, where Iceland and a few other hotspots show low-velocity anomalies (1% slower than the global average) at the depth between 100 and 200 km, making them distinctly different from anomalies under ridges. This difference suggests that ridges and hotspots may have intrinsically different driving mechanisms.

We discuss the relationship between S-wave velocity under ridges and spreading rates, and then interactions between ridges and plumes. The former gives further evidence that ridges are associated with shallow, passive upwelling. The passive upwelling, we refer to here is an adiabatic upwelling which fills the gap created by two separating plates. In this end-member view, plate spreading occurs by remote forces such as slab pull.

The S-wave velocity patterns beneath Afar and the South Atlantic near St Helena indicate lateral transport of material from hotspots to ridges. Such lateral transport has been invoked previously to explain complicated tectonic patterns and geochemical anomalies^{15–17}, but our observations now provide direct seismological evidence.

Ridges and hotspots

Figure 1 shows S-wave velocity anomalies underlying three ridge systems: the East Pacific Rise, the Pacific–Antarctica Ridge and the Southeast Indian Ocean Ridge. To remove possible local effects and errors, the velocity profiles have been averaged over 30° along the ridge axes. Each cross-section is parallel to the direction of spreading at that ridge. The hatched areas in Fig. 2, except the one in the Atlantic Ocean, show the corresponding locations of these stacked areas.

The figures show that all ridges are underlain by broad low-velocity regions and have very slow velocity anomalies in the upper 100 km. Regions with dotted patterns near the surface have velocities lower by more than 1% of the global average. The lowest velocities are found at depths shallower than 50 km in all cases. Regardless of spreading rate (~ 15 cm yr^{-1} , 8 cm yr^{-1} and 7 cm yr^{-1} from top to bottom), the depth extents of the -1% regions are ~ 100 km in all three cases. This consistency may reflect the limited depth resolution of the long-period surface wave data set. If shorter-period (< 75 -s) surface wave data were added, differences within the upper 100 km among the three cases might be revealed. Nonetheless, it is clear that the low-velocity anomalies are very shallow.

On the other hand, the horizontal extents of the -1% slow regions are $\sim 48^\circ$, 29° and 16° ($\sim 5,300$ – $1,800$ km) for the East Pacific Rise, Pacific–Antarctic Ridge and Southeast Indian

FIG. 2 Hatched areas are the locations for Figs 1 and 7. The lines AA', BB', CC' and DD' are parts of great circles that pass Hawaii, Iceland, Azores and Tristan da Cunha hotspots in Fig. 3a, respectively. The hotspot locations given by Morgan¹⁵ are aligned at centre. Two lines EE' and FF' show the locations for Fig. 5. The length of every curve is 60° (equal distance). The stars indicate the 46 locations in Fig. 4.

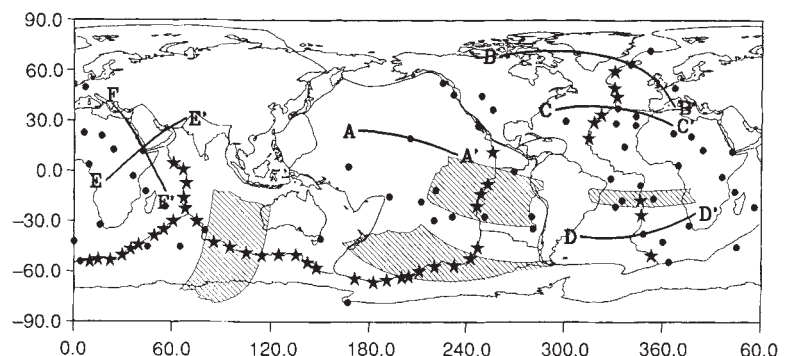
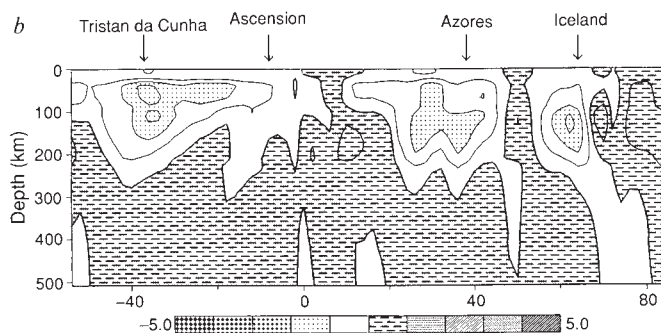
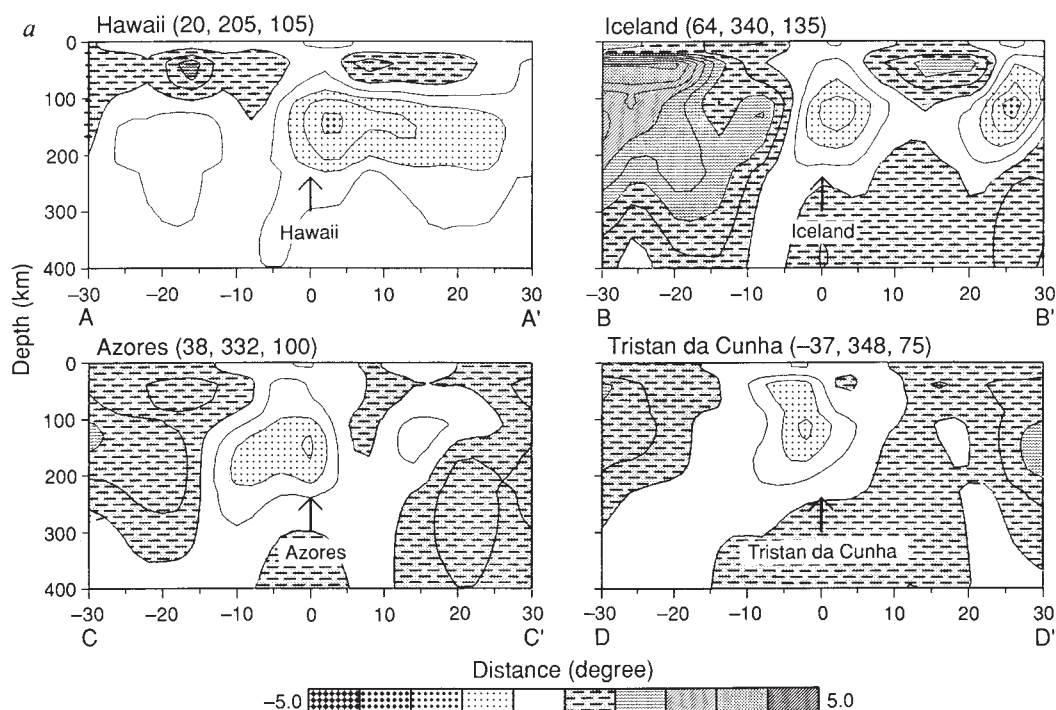


FIG. 3 *a*, Depth slices at four hotspot locations. The abscissa shows the great-circle distance from the hotspot in degrees. Numbers above each plot indicate (latitude, longitude, azimuth of great circle) at the hotspot. Note that low-velocity anomalies under hotspots are at ~100–200 km, deeper than low-velocity anomalies under ridges (Fig. 1). *b*, Depth slice along the Mid-Atlantic Ridge from south (left) to north (right). Three hotspots on or close to the ridge axis are associated with deep low-velocity anomalies, making them distinct from shallow anomalies under ridges (Fig. 1). This is unequivocal evidence that mechanisms under hotspots are different from those under ridges.



Ocean Ridge, respectively. This is correlated with spreading rate. It can be explained by large-scale flow, induced by plate motions which will extend thermal anomalies in the spreading direction: plates may entrain or drag the shallow mantle away from ridges. The correlation should be close to linear for such a model, and the data seem to suggest this, although Fig. 4 indicates some deviation from linearity. The deviation is probably caused by the uncertainty in the location of contours.

The velocity patterns under ridges are generally asymmetric below 100 km depth. Love wave dispersion data, which formed part of the data that led to the construction of the S-wave velocity model, showed asymmetry across some ridges⁷. The only region in which such asymmetry was clearly resolved seemed to be the South Atlantic, where it may be interpreted as high-temperature effects from hotspots on the eastern side of the ridge. Asymmetry in Fig. 1 may also reflect deep asymmetric thermal structure due to hotspots or other unknown causes that are unrelated to ridges, but these features are not well resolved.

In contrast to ridges, hotspots have minimum underlying mantle S-wave velocities at greater depths. Figure 3a shows vertical cross-sections of the S-wave velocity model near four main hotspots: Hawaii, Iceland, Azores and Tristan da Cunha. The corresponding locations are shown in Fig. 2. The lowest velocity anomalies (the dotted patterns under hotspots) are at depths between 100 and 200 km. Tristan da Cunha has a shall-

lower low-velocity anomaly, but this is still deeper than the low-velocity regions under ridges in Fig. 1.

The presence of low-velocity anomalies under hotspots at depths of 100–200 km implies a hotter or more volatile-rich mantle^{4,19} than that under ridges. These results seem consistent with the mantle plume hypothesis^{2,4,20}. A plume conduit of radius ~100–200 km rises from the deep mantle, spreads out radially beneath the lithosphere and produces a mushroom-like head (in terms of isotherms) with a diameter of ~1,000–2,000 km. The velocity anomalies under the hotspots may be related to the top of such plumes. Plume conduits are presumably narrower than our resolution, thus making it impossible to detect them in the model. Naturally the source depths of these plumes cannot be constrained from our model.

One question yet to be answered is whether the anomalies are elongated in the direction of plate motion. Such features are not clear from our model, if they even exist. For example, the low-velocity anomaly under Hawaii is stronger on the upstream (southeast) side than on the downstream (northwest) side of plate motion. Woods *et al.*¹⁸ also find normal surface-wave velocity along the Hawaii sea-mountain chain, which is consistent with the present result. Low-velocity anomalies in the direction of plate motion may exist on a smaller scale, but it seems that they cannot be very marked.

The 'head' of the plume as we use the term here is different from the initial head that Griffiths and Campbell^{21,22} invoke to explain continental flood basalts. Our head is the elongated thermal anomaly that always exists at the top of the plumes because of horizontally outward flow⁴. The 'head' in refs 21, 22 should move away from the plume, following the plate on top.

Other hypotheses exist for hotspots. Some speculate that hotspots occur at stagnation points of mantle return flow^{23–25}, others find that hotspots are located in areas with extensional stress in the upper mantle²⁶. We cannot discriminate between the varied interpretations, because our results lack sufficient resolution. It seems, however, that the model of a deep mantle plume is the most natural explanation for our images as well as for geoid and topography data²⁷.

The low-velocity anomalies under mid-ocean ridges indicate that mid-ocean ridges are shallow features, probably due to passive upwellings. Because density anomalies in the mantle are the driving force for active upwelling mechanisms (such as convective upwellings), active upwellings should produce low-velocity anomalies that extend much deeper than those observed under ridges. One of the best regions in which to examine such dichotomy is the Mid-Atlantic Ridge, because several hotspots exist on or very close to the ridge. Figure 3*b* shows the velocity profile along the Mid-Atlantic Ridge, and demonstrates that deep velocity anomalies exist under three hotspots. Shallow low-velocity anomalies under the ridge are not apparent, presumably because of the slow spreading rate of the Mid-Atlantic Ridge. This profile is critical to our argument, because without this evidence one can still argue that both ridges and hotspots are caused by active upwelling, the only difference being that upwellings reach close to the surface under ridges, whereas they are stopped below the lithosphere under hotspots. Shallow low-velocity anomalies under ridges can be intensified by decomposition melting effects. Figure 3*b* demonstrates that this view does not hold, because it cannot explain the deep velocity anomalies, extending to 200 km, under Iceland, Azores and Tristan da Cunha. These deep anomalies under the ridge axis can only be explained if hotspots are caused by a completely different mechanism from ridges.

A possible practical application of this is that the depth of low-velocity anomalies underlying ridges and hotspots could be

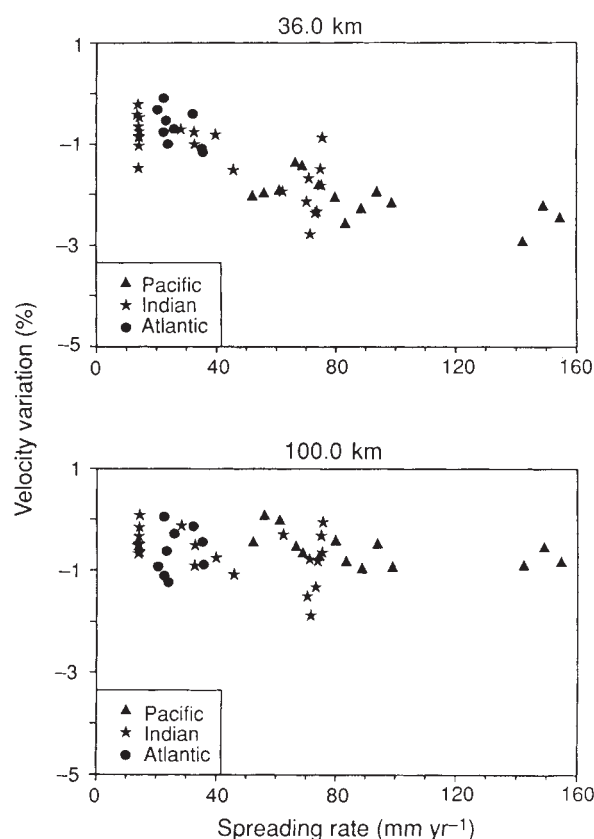


FIG. 4 S-wave velocity anomalies plotted against spreading rates under ridges. Forty-six locations selected from various ridges are shown in Fig. 2. Triangles, stars and solid circles indicate data from the Pacific, Indian and Atlantic oceans, respectively. Possible contamination from hotspots and triple-junctions is avoided. A strong correlation appears at a shallow depth (36 km), diminishes with depth and almost disappears by 100 km. This strongly suggests a predominantly surface-controlled process or a passive upwelling mechanism at ridges.

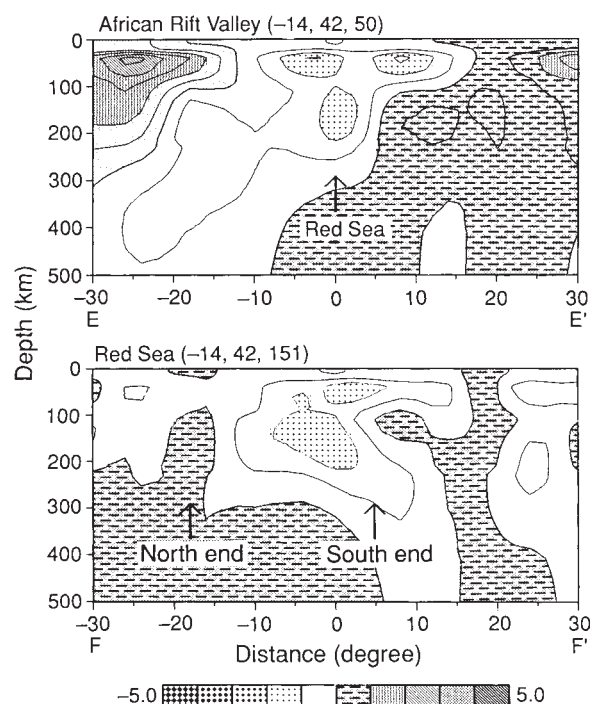


FIG. 5 Two depth slices in northeast Africa and Arabia. The slice in the lower panel goes through the Red Sea, and that in the top panel is perpendicular to it. The north and south end of the Red Sea are given in the bottom figure, which shows a deep velocity anomaly under the Red Sea. The Afar hotspot may be laterally feeding material to this region. The left of the figure shows a large low-velocity anomaly under Afar, which can be traced to a deep region under central Africa.

a useful criterion for determining whether surface features are caused by passive or active upwelling.

Passive upwellings at ridges

Figure 4 shows S-wave velocity anomalies at depth of 36 and 100 km, plotted against spreading rates under ridges. We selected forty-six points on ridge axes from the Pacific, Indian and Atlantic Oceans, based on spreading rate and spatial distribution. Possible contamination from hotspots, triple junctions and fracture zones were avoided. The locations of these points are denoted by stars in Fig. 2. Each symbol in Fig. 4 represents values averaged over 10° along the ridge axis. The spreading rates at all locations were calculated using the NUVEL-1 model²⁸.

The most notable feature in Fig. 4 is a correlation between S-wave velocity anomalies and spreading rates. The correlation is clearly stronger at 36 km depth. It then diminishes with depth and almost disappears by 100 km. This is further evidence that passive upwelling is associated with mid-ocean ridges. In the passive mechanism, as plates spread apart, material must well up at the spreading axes. A greater fraction of partial melting due to decompression occurs at shallower depth, which in turn produces slower S-wave velocity. As the fraction of melt is controlled by spreading rates, the correlation becomes stronger for shallower depths.

In the passive mechanism, there can be a broad, slow upwelling under ridges extending to substantial depths, but with less pronounced seismic anomaly. Material that wells up to fill the gap created by separating plates must eventually come from the deep interior. In our S-wave velocity results, there are low-velocity regions under ridges from the surface to ~ 300 km, showing that ridges overlie regions of generally slow velocity. The velocity varies much less, however, below 100 km. The only evidence for a broad, deep upwelling under a ridge can be found

under the East Pacific Rise, where a 0.5% slow velocity contour extends downwards to more than 200 km in depth (Fig. 1). Although this anomaly is only marginally significant statistically, this feature is consistent with the fact that the East Pacific Rise has the fastest spreading rate on Earth, and should be accompanied by the most extensive and deepest induced flow below the ridge.

Ridge (rift)–plume interaction

The first example for the ridge–hotspot interaction is in Fig. 5, which shows two depth slices, one along the Red Sea (FF' in Fig. 2) and the other (EE') perpendicular to it. The mid-points of both slices are slightly north of the Afar hotspot. The slice in the lower panel is along the Red Sea and shows a low-velocity anomaly that is bounded at the north and south ends of the Red Sea and extends to a depth of ~250 km. The length of the Red Sea is more than 20°, so resolution is not a problem in this map. The lowest-velocity anomaly below 100 km depth is less than -1%, a significant anomaly greater than two standard deviations. Note that the depth of the lowest-velocity anomaly under the Red Sea is between 100 and 200 km, similar to velocity anomalies under hotspots. Furthermore, this low-velocity anomaly seems to be connected to a deeper source at the southern end, located near the Afar hotspot. The shapes of the contours at 0 and 0.5% suggest that material may be fed laterally or obliquely from this hotspot to rifts in the Red Sea. This interpretation, however, may not be robust statistically, because the shape of the 0% contour is not reliable.

The perpendicular slice in the top panel shows extension of the low-velocity anomaly towards central Africa. This figure suggests that there is a broad upwelling of hot materials from central Africa toward the Red Sea region. This may seem to be similar to the model proposed by Schilling *et al.*²⁹ on the basis of geochemical data. There is, however, a difference in the size of the oblique upwelling. The S-wave velocity anomaly shown here requires a much broader upwelling than the single plume with a size of ~100–300 km proposed in ref. 29. The S-wave velocity anomaly from such a narrow plume will be less than 0.5% on our resolution scale of 1,000 km. Either multiple plumes or a very broad hot upwelling are required to explain the low-velocity anomaly in Fig. 5. One interpretation is that a few hotspots under central Africa are contributing to this broad anomaly. The number of hotspots in central Africa varies from one list to another^{30–32}, but could be substantial.

Another interesting aspect of this anomaly is its obliqueness. In numerical simulations of mantle convection, it is often found that plumes go around cold temperature anomalies and come up obliquely. As there are huge continental shields in Africa, characterized by cold temperatures and fast seismic velocities, which are presumably being eroded by these plumes' activities, it may not be surprising to find such oblique plumes that have avoided cold continental shields. Alternatively, the plume may have come up straight and the head been dragged in the northeast

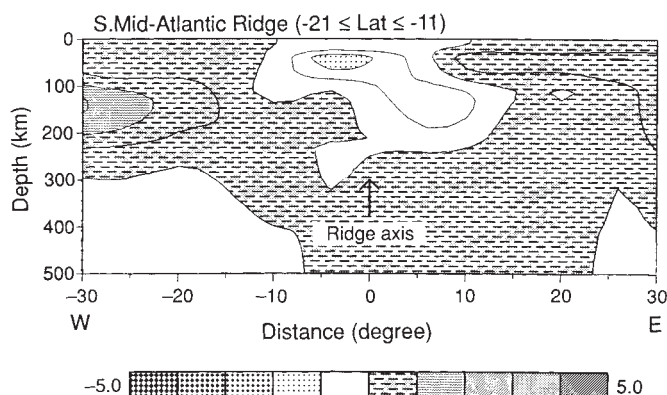


FIG. 6 Depth slice showing averaged S-wave velocity, perpendicular to the ridge axis near St Helena in the South Atlantic. The low-velocity anomaly at the ridge is elongated and deepens towards the east. Because the hotspot at St Helena is near its east end, the shape of this low-velocity anomaly suggests the existence of a channel which transports material from the hotspot to the ridge axis.

direction, while weakening the region under the Gulf of Aden. From the geochemical data²⁹ and past reconstruction of hotspot tracks^{31,32}, the latter hypothesis has sometimes been preferred. But there is no strong reason to adhere to the idea of a vertically straight plume, when there is ample evidence of temperature anomalies in the upper mantle.

Another example of ridge–hotspot interaction can be found in the Southern Atlantic. Figure 6 shows an averaged vertical section perpendicular to the ridge axis. East–west slices were averaged over 10° along the ridge axis near St Helena. The ridge axis is aligned along the centre, denoted by the arrow. It should be noted that the low-velocity anomaly under the ridge is elongated and deepens toward the east. St Helena lies at the far eastern end of this velocity anomaly. Geochemical evidence^{16,17} suggests that there is transport of material from the hotspot to the nearby ridge. This low-velocity anomaly may thus be interpreted as evidence for the channel that transports material from the hotspot to the ridge. Although seismic evidence alone may not be robust because of its low resolution, combined geochemical and geophysical data strongly support the idea that hotspots feed material to nearby ridges from a distance in the South Atlantic. This kind of argument has been invoked to explain the complicated tectonics near Iceland^{33,34}.

We have described two examples of ridge–hotspot interactions. Although this kind of interaction may be common, our resolution limit prevents our observing smaller-scale interactions or more details of mechanisms at these two locations. Higher-resolution seismic results as well as geochemical studies are needed to clarify how these interactions occur. □

Received 19 August; accepted 29 November 1991.

1. Zindler, A. & Hart, S. *Ann. Rev. Earth planet. Sci.* **14**, 493–571 (1986).
2. Morgan, W. J. *Geol. Soc. Am. Mem.* **132**, 7–22 (1972).
3. Spiegelman, M. & McKenzie, D. P. *Earth planet. Sci. Lett.* **83**, 137–152 (1987).
4. White, R. S. & McKenzie, D. J. *J. geophys. Res.* **94**, 7685–7729 (1989).
5. Scott, D. R. & Stevenson, D. J. *J. geophys. Res.* **94**, 2973–2988 (1989).
6. Kent, G. M., Harding, A. J. & Orcutt, J. A. *Nature* **344**, 650–653 (1990).
7. Zhang, Y.-S. & Tanimoto, T. *Phys. Earth planet. Int.* **66**, 160–202 (1991).
8. Zhang, Y.-S. thesis, California Institute of Technology (1991).
9. Nataf, H. C., Nakanishi, I. & Anderson, D. L. *J. geophys. Res.* **91**, 7261–7307 (1986).
10. Woodhouse, J. H. & Dziewonski, A. M. *J. geophys. Res.* **89**, 5953–5986 (1984).
11. Tanimoto, T. & Anderson, D. L. *Geophys. Res. Lett.* **11**, 287–290 (1984).
12. Tanimoto, T. *Geophys. J. R. astr. Soc.* **93**, 321–334 (1988).
13. Wong, Y.-K. thesis, Harvard Univ. (1989).
14. Tanimoto, T. *Geophys. J. Int.* **100**, 327–336 (1990).
15. Morgan, W. J. *J. geophys. Res.* **83**, 5355–5360 (1978).
16. Schilling, J. G. *et al. Nature* **313**, 187–191 (1985).
17. Schilling, J. G. *Nature* **314**, 62–67 (1985).
18. Woods, M. T., Leveque, J.-J., Okal, E. A. & Cara, M. *Geophys. Res. Lett.* **18**, 105–108 (1991).

19. Bonatti, E. *Science* **250**, 107–111 (1990).
20. Duncan, R. A. & Richards, M. A. *Rev. Geophys.* **29**, 31–50 (1991).
21. Griffiths, R. W. & Campbell, I. H. *Earth planet. Sci. Lett.* **99**, 66–78 (1990).
22. Campbell, I. H. & Griffiths, R. W. *Earth planet. Sci. Lett.* **99**, 79–93 (1990).
23. Chase, C. G. *Geophys. J. R. astr. Soc.* **56**, 1–18 (1979).
24. Hager, B. H. & O'Connell, R. J. *J. geophys. Res.* **86**, 4843–4867 (1981).
25. Ricard, Y., Vigny, C. & Froidevaux, C. *J. geophys. Res.* **94**, 13739–13754 (1989).
26. Liu, H.-S. *Tectonophysics* **65**, 225–244 (1980).
27. Sleep, N. H. *J. geophys. Res.* **95**, 6715–6736 (1990).
28. DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. *Geophys. J. Int.* **101**, 425–478 (1990).
29. Schilling, J.-G., Kingsley, R. H., Hanan, B. B. & McCully, B. L. *J. geophys. Res.* (submitted).
30. Wilson, J. T. *Tectonophysics* **19**, 149–164 (1973).
31. Vogt, P. R. *J. geophys. Res.* **86**, 950–960 (1981).
32. Morgan, W. J. *The Sea* **7**, 443–487 (Wiley, New York, 1981).
33. Vink, G. E. *J. geophys. Res.* **89**, 9949–9959 (1984).
34. Vink, G. E., Morgan, W. J. & Vogt, P. R. *Scient. Am.* **50–57** (April 1985).

ACKNOWLEDGEMENTS. We thank D. L. Anderson, D. Dreger, B. L. N. Kennett, R. Jeanloz, M. Kohler, T. Lay and N. Sleep for comments. This research was supported by the NSF.

Postsynaptic contribution to long-term potentiation revealed by the analysis of miniature synaptic currents

Toshiya Manabe, Pius Renner & Roger A. Nicoll*

Departments of Pharmacology and Physiology, University of California, San Francisco, California 94143-0450, USA

Miniature excitatory synaptic currents were recorded from CA1 pyramidal cells in hippocampal slices to study the site of the persistent change in synaptic efficacy during long-term potentiation. Induction of long-term potentiation produced a large increase in the amplitude of these currents. Such a change in amplitude suggests an increase in postsynaptic transmitter sensitivity.

HIGH-FREQUENCY activation of certain synapses causes long-lasting enhancement of synaptic transmission¹⁻³. This long-term potentiation (LTP) is the main candidate for a cellular mechanism for learning and memory in the vertebrate central nervous system (CNS). LTP is initiated postsynaptically in the CA1 region of the hippocampus⁴⁻⁷, but controversy remains over the site responsible for the persistent change. There is evidence for involvement of both the presynaptic terminal and the postsynaptic cell⁸⁻¹⁸.

Quantal analysis, which was originally used to study the mechanism involved in synaptic transmitter release at the neuromuscular junction, is based on the occurrence of amplitude fluctuations of synaptic transmission. It was proposed that transmitter is released in multimolecular packets of constant size, thought to correspond to the contents of single vesicles in the presynaptic terminal, and that evoked release is made up of a number of simultaneously released vesicles¹⁹. This analysis provides a powerful technique for determining the site of the synapse at which changes in the strength of transmission occur^{19,20}. A change in the number of quanta released with each stimulus indicates a presynaptic mechanism, whereas a change in the size of quantal events signifies a change in the postsynaptic response to neurotransmitter. Using this technique, it has been proposed that the site of the persistent change during LTP is primarily presynaptic (refs 12-15, but see ref. 16). But there remain questions about how directly the model of the neuromuscular junction can be applied to synapses of the CNS, as many of the assumptions on which quantal analysis is based have not yet been proved to hold for CNS synapses^{17,18}.

At the vertebrate neuromuscular junction, the validity of quantal analysis has been greatly strengthened by the study of spontaneous miniature synaptic events. Analysis of the amplitude of miniature synaptic events has provided a simple and direct method for monitoring postsynaptic transmitter sensitivity and is free of many of the assumptions associated with quantal analysis of evoked release²¹. The use of miniature excitatory postsynaptic currents (m.e.p.s.cs) or potentials in the CNS in analysing synaptic transmission has been limited. First, unlike vertebrate muscle fibres, CNS neurons are multiply innervated and therefore the synaptic origin of a given m.e.p.s.c. is unknown. Second, spontaneous firing of presynaptic neurons can contribute events containing multiple quanta. Third, the size of m.e.p.s.cs is at least an order of magnitude smaller than

their counterpart at the neuromuscular junction, making their detection difficult. Here we have adopted strategies to minimize these problems so that we could adequately test whether quantal size changes during LTP.

Properties of spontaneous e.p.s.cs

We used whole-cell recording from CA1 pyramidal cells in hippocampal slices; this greatly improves the signal-to-noise ratio compared with conventional recording techniques, permitting routine recording of spontaneous inward currents (Fig. 1a). These currents, as well as the electrically evoked e.p.s.cs, were entirely blocked by the non-NMDA glutamate receptor antagonist, 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) (data not shown), which indicates that these inward currents are spontaneous e.p.s.cs. An example of the averaged spontaneous e.p.s.c. obtained during the control period in Fig. 1a is shown in Fig. 1c (inset). The average amplitude for these events recorded at -90 mV ranged from 4 to 15 pA in different cells, being about 5 pA in Fig. 1. A plot of an amplitude histogram (Fig. 1c, open histogram) revealed that the distribution is skewed toward larger events, as reported earlier²². It is important to emphasize that, even with whole-cell recording, the detection of small events was limited by the signal-to-noise ratio. The noise distribution (Fig. 1c, filled histogram) indicates that some small events were below the detection threshold. Throughout this study we have used cumulative plots of spontaneous e.p.s.c. amplitudes. In these plots the ordinate represents cumulative relative frequency of spontaneous e.p.s.cs and the abscissa, their amplitude. The cumulative relative frequency is the rank of events normalized by the total number when they are sorted in ascending order of their amplitude. This method of display has advantages over standard histograms because it allows easier comparison of two sets of data and the display of every data point without arbitrary binning²³ (Fig. 1d). The frequency of these events (typically about 1 Hz) varied considerably from cell to cell and during recordings from individual cells, making quantitative analysis difficult. Tetrodotoxin (TTX), which blocks action potential-evoked release, in most cases (five out of six cells), had little effect on the frequency of spontaneous events in the absence of nerve stimulation (mean change = $4 \pm 15\%$, $n = 6$). This finding indicates that in our conditions most detected spontaneous events are true m.e.p.s.cs. The cumulative plots of amplitude distributions were similar before and after application of TTX (Fig. 1d, right panel).

At the neuromuscular junction, synaptic activation increases the rate of spontaneous synaptic events^{24,25}. A similar phenomenon occurs at hippocampal synapses. Thus low-frequency electrical stimulation of Schaffer collateral/commisural afferents (0.2 Hz with paired stimuli separated by 50 ms) resulted in a clear increase in the frequency of events (Fig. 1a, b). The increase in frequency was typically around twofold. The cumulative amplitude distribution of these events during stimulation was similar to that obtained in the absence of stimulation (Fig. 1d, left panel). Application of TTX in this experiment, which blocked evoked responses, lowered the frequency to control values (Fig. 1b). Although changing the frequency of

* To whom correspondence should be addressed.

spontaneous events had little effect on the amplitude distribution, an increase in the size of these events could easily be detected by increasing the driving force for the e.p.s.c. Figure 1e shows the results from one of seven experiments, demonstrating that the amplitude distribution of spontaneous e.p.s.c.s shifted to the right with hyperpolarization.

Synaptic activation of a subset of synapses doubles the frequency of spontaneous e.p.s.c.s (Fig. 1a–d), suggesting that many of the spontaneous synaptic events recorded during repetitive synaptic activation originate from the synapses being stimulated. Thus spontaneous synaptic events can be recorded that are biased toward a small subset of synapses that can then be manipulated by stimulation to give LTP.

LTP increases spontaneous e.p.s.c. amplitude

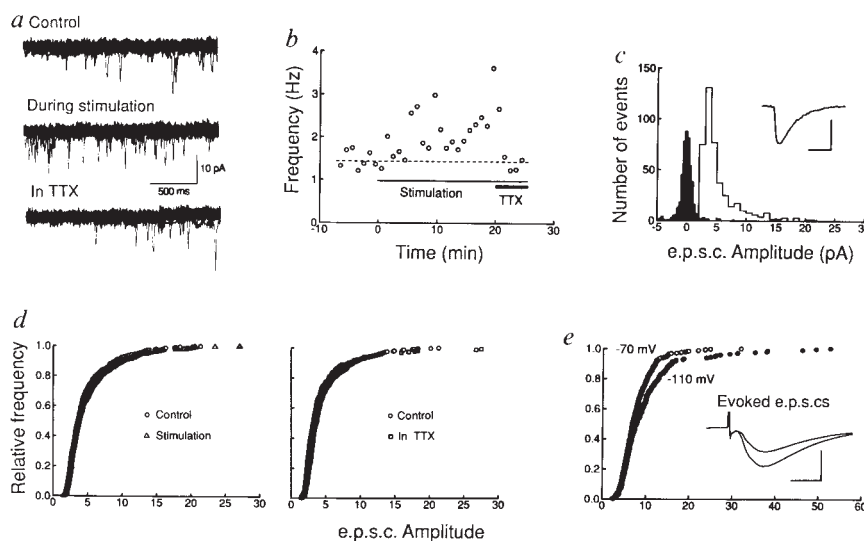
We next analysed the effect of LTP on the amplitude of spontaneous e.p.s.c.s collected as described above. Figure 2a shows the averaged stimulus-evoked e.p.s.c.s and the averaged spontaneous e.p.s.c.s recorded over the same period from one cell. Sample spontaneous e.p.s.c.s are shown in Fig. 2b and the cumulative plots of these events are shown in Fig. 2c. LTP was

induced by depolarizing the cell during synaptic stimulation (pairing), which was always continued at constant strength and frequency throughout the experiment. After the pairing the stimulus-evoked e.p.s.c. was considerably increased, indicating that LTP had occurred at these synapses (Fig. 2a, right panel). An increase in the size of the spontaneous e.p.s.c.s also occurred, as shown in the sample traces (Fig. 2b), the averaged spontaneous e.p.s.c. (Fig. 2a, right panel) and the cumulative plots (Fig. 2c). In six of the seven cells in which LTP occurred in the evoked e.p.s.c., the difference in the amplitude distribution of the spontaneous e.p.s.c.s before and after LTP was significant at the 0.05 level (Kolmogorov–Smirnov test, see the legend to Fig. 1). The mean amplitude of the spontaneous e.p.s.c.s, measured 15 to 20 min after pairing, increased by $35 \pm 8\%$ ($n = 7$). The increase in the mean amplitude of the evoked e.p.s.c. was $76 \pm 8\%$. A difference in the magnitude of the effect of LTP on the evoked and spontaneous e.p.s.c.s can be explained by the fact that not all of the spontaneous e.p.s.c.s came from those synapses activated by electrical stimulation.

In five cells we have compared the time course of the change in spontaneous e.p.s.c. size with the change in evoked e.p.s.c.

FIG. 1 The effect of synaptic stimulation on spontaneous e.p.s.c.s. *a*, Series of 10 superimposed traces illustrating spontaneous e.p.s.c.s before (control), during 0.2 Hz electrical stimulation of Schaffer collateral/commissural afferents (during stimulation), and after the addition of $1 \mu\text{M}$ TTX (in TTX). *b*, Graph of time course of the change in frequency of spontaneous e.p.s.c.s by electrical stimulation. *c*, Histogram of the size distribution of control spontaneous e.p.s.c.s (open histogram) and the distribution of the baseline noise (solid histogram) measured in the same way. The inset shows an average of 40 spontaneous e.p.s.c.s recorded during the control period. The calibration is 5 pA and 20 ms. *d*, Cumulative plots before and during 0.2 Hz stimulation (left) and after blockade of action potentials with TTX (right). *e*, Effect of membrane hyperpolarization from -70 to -110 mV on the cumulative plots of the spontaneous e.p.s.c. amplitudes. The evoked e.p.s.c.s in the same cell are shown in the inset. The calibration is 50 pA and 10 ms.

METHODS. Transverse hippocampal slices (400–500 μm) from 3–5-week-old guinea-pigs were cut with a vibratome in ice-cold superfusing medium. The slices were placed in a holding chamber for ≥ 1 h and then transferred, as needed, to the recording chamber. The superfusing medium, unless otherwise stated, contained (in mM): NaCl, 119; KCl, 2.5; MgCl_2 , 1.3; CaCl_2 , 2.5; NaHCO_3 , 26.2; NaH_2PO_4 , 1; glucose, 11; picrotoxin 0.1 and was bubbled with 95% O_2 /5% CO_2 (pH 7.4). Picrotoxin was present in all experiments to eliminate inhibitory synaptic events. In one set of experiments a superfusing medium containing 10 mM Ca^{2+} was applied for 2 min with NMDA. This medium lacked NaH_2PO_4 and NaHCO_3 was reduced to 10 mM. In all experiments surgical cuts isolated the CA1 region from all other regions of the hippocampus (see Fig. 4a). The e.p.s.c.s were recorded with a patch electrode (3–7 M Ω resistance) in the whole-cell voltage clamp mode (Axopatch 1D) using the 'blind' approach^{33,34}. The series resistance and input resistance were monitored throughout the experiment and if either one changed by more than 15%, the experiment was rejected. If the series resistance changed during the experiment, it usually increased, which would reduce the size and slow the time course of the recorded events. Series resistance compensation was not used to maintain as high a signal-to-noise ratio as possible. The pipette solution, unless otherwise stated, contained (in mM): caesium gluconate, 122.5; CsCl, 17.5; HEPES, 10; EGTA, 0.2; NaCl, 8; Mg-ATP, 2; GTP, 0.3. The pH was adjusted to 7.2. Osmolarity of the solution, which was routinely measured, was 295–300 mOsm. Unless otherwise stated, e.p.s.c.s were recorded at -90 mV. All membrane potential values were corrected for a liquid junction potential of 10 mV. Records were filtered at 1 kHz, digitized at 3–5 kHz on a TL-1 interface (Axon) or stored on magnetic tape then transferred to a computer. To examine the effect of LTP on spontaneous e.p.s.c.s, excitatory afferent fibres were stimulated close to the recording electrodes with a glass pipette filled with superfusing solution. Large stimulus strengths were used to ensure activation of a large number of synapses. LTP was induced by pairing



depolarization (to -10 mV) with 30–40 paired-pulse stimulations (total of 60–80 stimuli) without changing stimulus frequency (0.2 Hz). All values are expressed in terms of mean $\pm$ s.e.m. To determine whether the amplitude distribution of two populations of spontaneous e.p.s.c.s differed from each other, the nonparametric Kolmogorov–Smirnov test was used²³ instead of, for example, Student's *t*-test, because the amplitude distributions deviated considerably from a normal distribution. The amplitude of spontaneous e.p.s.c.s was measured by a computer program for automatic detection. Threshold detection level was between 2–6 pA. With lower series resistance, the size of both the noise and the signal was relatively larger, thus requiring a higher detection threshold. The selected threshold level remained constant for the entire analysis of the experiment. A threshold level in each experiment was selected so that about 20% of events were rejected as noise. In some experiments CNQX was added at the end of the experiment to verify that virtually all of the events that reached threshold and had the wave form of an e.p.s.c. were blocked. All events larger than this level were included if they had rise times < 7 ms, had the general shape of an e.p.s.c., and did not arise from the decaying phase of a previous event. Although these events arising from the decaying phase of a previous event were not included in the amplitude measurements, they were included in the frequency measurements. A rise time of < 7 ms was used to exclude events generated at distal dendritic sites. To measure the amplitude of spontaneous e.p.s.c.s the computer calculated the difference between the average of 2 ms before the onset of the e.p.s.c.s and the average of 2 ms around the peak. For averaging spontaneous e.p.s.c.s, all events lacking superimpositions were included. Most experiments in each of the three categories were reanalysed in a blind fashion to verify the conclusions of the analysis. The drugs used were TTX (Sankyo), CNQX (CRB), APV (CRB), BAPTA (Molecular Probes), NMDA (Sigma), kainate (Sigma) and picrotoxin (Sigma).

size after inducing LTP. One example is shown in Fig. 3, which indicates that, as with the evoked e.p.s.c. (Fig. 3c), the potentiation in spontaneous e.p.s.c.s occurs abruptly after the pairing and lasts for at least 40 min (Fig. 3a). In addition, the potentiation is not associated with any consistent change in the frequency of spontaneous e.p.s.c.s (Fig. 3b). For all of the experiments in which the evoked e.p.s.c. was potentiated ($n = 7$) the average frequency of spontaneous e.p.s.c.s before LTP was 2.4 ± 0.4 Hz and after LTP was 2.7 ± 0.3 Hz. The effect of LTP on the recorded frequency was dependent on the detection threshold. For example, in Fig. 3b a threshold of 2 pA was used and little change in frequency was recorded associated with an increase in mean amplitude of 38%. Little change in frequency was detected for thresholds below 4.5 pA, whereas with larger thresholds a clear increase in recorded frequency was observed. This suggests that at the lower thresholds most of the spontaneous e.p.s.c.s that undergo LTP were detected before LTP. In contrast, the increase in mean amplitude remained constant when different thresholds (2–6 pA) were used.

In three cells, one of which was recorded in the presence of the NMDA receptor antagonist, D-2-amino-5-phosphonovaleric acid (APV) (30 μ M), no LTP occurred in the evoked responses after pairing, and there was no change in the size of the spontaneous e.p.s.c.s in any of these cells (mean = $98 \pm 9\%$ of control). Thus the changes in the distribution of spontaneous e.p.s.c. amplitude were specifically related to the occurrence of LTP in the evoked pathway.

Several possibilities can be considered in attempting to explain a shift in the amplitude distribution of the spontaneous e.p.s.c.s. First, LTP might cause a homogeneous increase in the frequency of all events regardless of size. This, however, is not compatible with the data, as such a change would not alter the

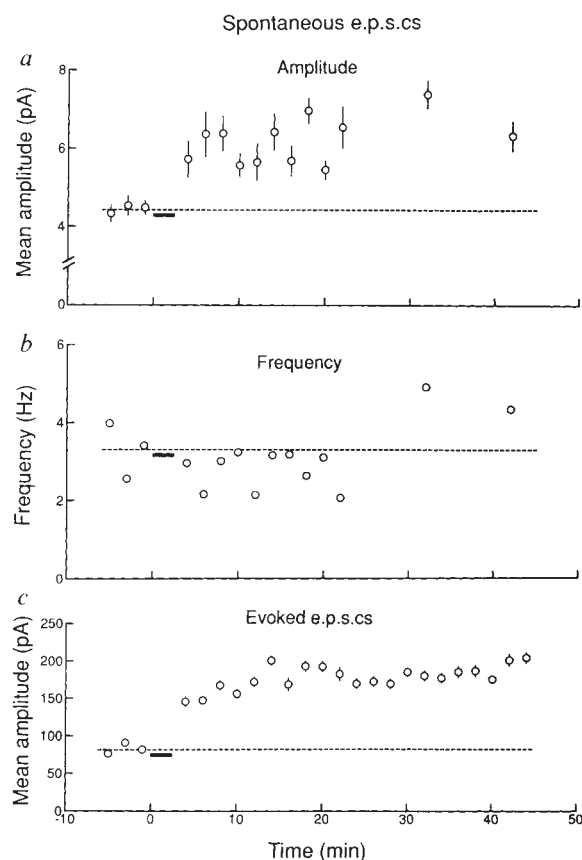


FIG. 3 Time course of the change in spontaneous e.p.s.c. amplitude elicited by LTP. The data are from the same cell as in Fig. 2. *a*, Effect of inducing LTP on the spontaneous e.p.s.c. amplitude. All spontaneous e.p.s.c.s larger than 2 pA were measured. Each point represents the average amplitude of 250–550 events over a 2 min period. *b*, LTP has little effect on the frequency of spontaneous e.p.s.c.s. *c*, Mean amplitude of evoked e.p.s.c.s over 2 min periods for the same experiment. Each point is the average of 24 events. LTP was induced by pairing for 2.5 min starting at time 0 (see bar).

shape of the amplitude distribution. Another possibility is that LTP might cause a selective increase in frequency of large events. But the small increase in e.p.s.c. frequency observed during LTP (13%) makes this explanation very unlikely. Even if we assume in Fig. 2 that only the frequency of events larger than 5 pA had been increased, the rate of these events would have had to increase 8.5-fold to account for the change in mean amplitude, and this would have increased the overall observed frequency roughly 3.5-fold. Finally, e.p.s.c.s of all sizes could become larger after LTP or growth could be limited to a select size class. Both of these possibilities are consistent with our results.

The above results indicate that during LTP the size of spontaneous e.p.s.c.s increases. But these experiments were done in the absence of TTX, which blocks action potential discharge, because afferent fibre stimulation was used to increase selectively spontaneous e.p.s.c. frequency. It is conceivable that some of the spontaneous e.p.s.c.s could have arisen from spontaneous action potential discharge, even though the CA1 region was surgically isolated from the rest of the hippocampus. Because resistance to TTX is essential for rigorously defining true miniature e.p.s.c.s, we devised two types of experiments in which the effects of LTP on the amplitude of m.e.p.s.c.s could be examined even in the presence of TTX.

LTP increases sucrose-evoked m.e.p.s.c. amplitude

In the first type of experiment, a pipette containing 0.5 M sucrose in superfusing medium was positioned in the slice close to the

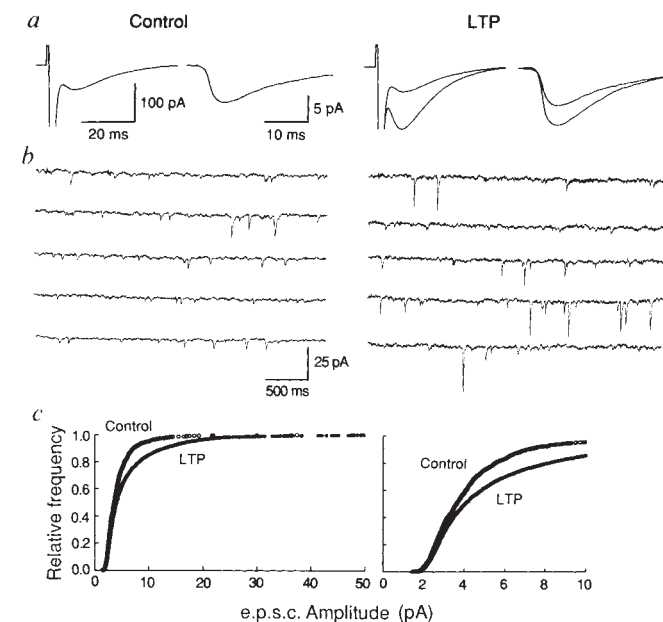
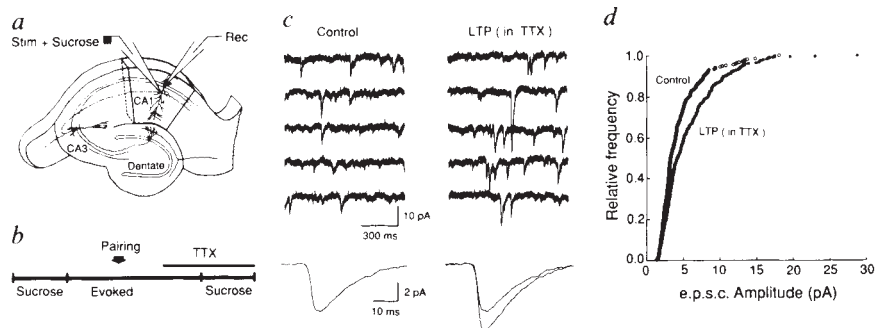


FIG. 2 Induction of LTP in the evoked e.p.s.c. also increases the amplitude of spontaneous e.p.s.c.s. *a*, Evoked e.p.s.c. (average of 20 traces) (left trace) and spontaneous e.p.s.c. (average of 100 events) (right trace) before (left panel) and 15 min after LTP (right panel, which also contains the averaged traces obtained before LTP). Spontaneous e.p.s.c.s greater than 3.5 pA were averaged. The large stimulus artefact is due to the proximity of the glass stimulating electrode to the recording electrode. *b*, Examples of individual spontaneous e.p.s.c.s obtained between six consecutive evoked e.p.s.c.s before (left) and 15 min (right) after LTP. *c*, Cumulative frequency distribution of spontaneous e.p.s.c. amplitude before and after LTP (includes all events > 2 pA). The control curve was constructed from events collected during a 6 min period before pairing. The curve for LTP was constructed from all events occurring for 40 min after pairing. The graph on the left is shown with the abscissa expanded on the right.

FIG. 4 Amplitude of m.e.p.s.cs evoked by sucrose is increased during LTP. *a*, Experimental arrangement and the cuts made to isolate the CA1 region. A pipette containing 0.5 M sucrose in superfusing solution was positioned near the recording electrode and was used both to evoke m.e.p.s.c. release and to electrically stimulate excitatory afferents. Sucrose was applied by brief (5–50 ms) pressure pulses (5 p.s.i.) (Picospritzer II), repeated every 5 to 30 s. *b* shows protocol for the experiment. *c*, Five consecutive traces before and after LTP (upper), and averages of 120 m.e.p.s.cs larger than 3 pA (lower). The averaged m.e.p.s.c. obtained after LTP (right) is superimposed on the control trace. *d*, Cumulative plots of amplitude distribution of m.e.p.s.cs larger than 2 pA before and 20 min after LTP. The data after LTP were collected in the presence of TTX.



recorded cell (Fig. 4*a*). At the neuromuscular junction, hypertonic solution increases the rate of spontaneous synaptic events²⁶. The protocol for this experiment is shown in Fig. 4*b*. The hypertonic solution was ejected for 2 min in the absence of electrical stimulation and the sucrose-evoked e.p.s.cs were recorded. Control e.p.s.cs were then evoked by strong focal electrical stimulation (0.2 Hz) through the same pipette in an attempt to include as many synapses in the vicinity of the sucrose pipette as possible. LTP was then induced by pairing with the evoked e.p.s.c., TTX was added to the superfusing medium and sucrose-evoked m.e.p.s.cs were again collected. The size of the sucrose-evoked e.p.s.cs, recorded in the presence of TTX, was increased after LTP (Fig. 4*c, d*). The average increase in m.e.p.s.c. amplitude was $15 \pm 4\%$ ($n = 7$), measured 15–20 min after pairing. In six of the seven cells in which LTP was observed in the evoked e.p.s.c. the potentiation of m.e.p.s.cs was significant at $P < 0.05$. Alternatively in five cells, two of which were bathed in APV, no LTP occurred in the evoked response after pairing and there was no significant change in the size of the sucrose-evoked m.e.p.s.cs (mean = $96 \pm 2\%$ of control). Thus the increase in size of m.e.p.s.cs is related to LTP in the evoked pathway and cannot be explained by a direct effect of the hypertonic solution, which has been reported to increase quantal size slowly at the frog neuromuscular junction²³.

NMDA increases m.e.p.s.c. amplitude

The second type of experiment took advantage of the observation that application of NMDA to slices results in a potentiation of evoked e.p.s.ps^{27,28}. This experiment has the advantage that all of the synapses in the vicinity of the application should be affected. In initial experiments NMDA was applied by pressure to the surface of the slice in the absence of TTX so that the effect on the evoked and spontaneous e.p.s.cs could be compared. An example of one of these experiments is shown in Fig. 5. Following a brief depression of the evoked response after NMDA application, the sizes both of the electrically evoked and of spontaneous e.p.s.cs increased. The time course of the development of the increase was similar for the two responses, although in this cell the increase in the spontaneous e.p.s.cs was more persistent. After NMDA application the histograms and cumulative plots of spontaneous e.p.s.cs show that the distribution of events was clearly shifted to the right, with a decrease in occurrence of small events (Fig. 6*a*). When the entire experiment was done in TTX, a similar shift was evident after application of NMDA (Figs 6*b* and 7), confirming that the amplitude of m.e.p.s.cs increased. The average potentiation of m.e.p.s.cs in experiments with the superfusing medium containing 4 mM Ca^{2+} was $60 \pm 8\%$ ($n = 8$). The potentiation of m.e.p.s.cs by NMDA could be elicited either by pressure application (10 out of 10 cells) (Figs 5 and 6*a*) or by bath application (9 out of 10 cells) ($10 \mu\text{M}$ for 2 min) (Figs 6*b* and 7). Bath application of

kainate ($10 \mu\text{M}$ for 2 min, $n = 3$) in TTX failed to potentiate the m.e.p.s.cs. In six of the experiments ($4 \text{ mM } \text{Ca}^{2+}$) using pressure application of NMDA, the evoked e.p.s.c. was also recorded and was potentiated by $67 \pm 25\%$. The close similarity in the degree of potentiation in the evoked and spontaneous e.p.s.cs suggests that the potentiation of the evoked e.p.s.c. by NMDA can be largely accounted for by the potentiation of the spontaneous e.p.s.cs.

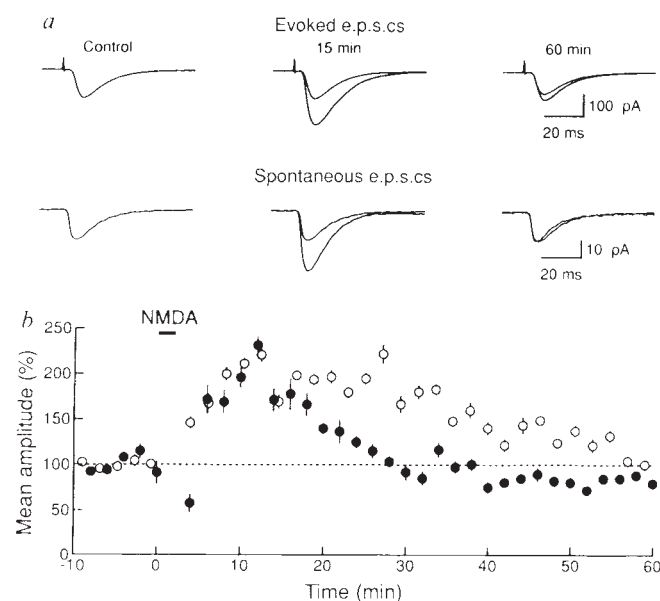


FIG. 5 NMDA application potentiates both the evoked and spontaneous e.p.s.cs. All data are from the same cell. *a*, Averages of 30 e.p.s.cs evoked by electrical stimulation before (control), 15 and 60 min after the application of NMDA (top) and averages of 100 spontaneous e.p.s.cs (including events larger than 6 pA) recorded during the same time period as the evoked e.p.s.cs (bottom). The control records are superimposed on the records obtained after the application of NMDA. *b*, Graph of the time course of the potentiation of the evoked (●) and spontaneous e.p.s.cs (○), partially illustrated in *a*. Each point for the spontaneous e.p.s.cs is the average of about 150–300 events recorded over 2 min periods. Each point for the evoked e.p.s.c. is the average of 24 events.

METHODS. A broken patch electrode filled with 500 μM NMDA in superfusing medium was positioned on the surface of the slice close to the recording electrode. The amplifier was switched from voltage clamp mode to current clamp mode during NMDA application. NMDA was applied by pressure ejection until the cell depolarized at least 30–40 mV and then the NMDA-containing pipette was withdrawn. The superfusing medium contained 4 mM Ca^{2+} . The internal solution contained potassium gluconate instead of caesium gluconate and KCl instead of CsCl.

At the peak of the NMDA-induced potentiation there was an increase in the recorded frequency of m.e.p.s.cs above the threshold set for detection ($42 \pm 18\%$, $n = 20$). Although we cannot exclude some real change in frequency of m.e.p.s.cs, which would indicate a presynaptic action, an alternative explanation for this change is the emergence of enlarged m.e.p.s.cs from below the threshold for detection. This would be a considerably greater issue with NMDA application than with the stimulus-induced LTP, because all of the synapses on the cell would be exposed to NMDA and the magnitude of the potentiation was much greater. A frequency change alone cannot explain the dramatic shift in the amplitude distributions of m.e.p.s.cs after the application of NMDA because (1) the emergence of large events was clearly associated with an absolute decrease in the number of small events recorded for the same duration in all cells examined and (2) large events, which were often entirely absent during the control period, were abundant after the application of NMDA (Fig. 6).

How similar is NMDA-induced potentiation to LTP? NMDA-receptor activation is required for the induction of LTP²⁷. NMDA application also causes a potentiation of synaptic transmission^{27,28}. This potentiation typically lasts for about 30 min^{27,28}, but has been reported to be longer lasting in a superfusing medium containing an elevated $\text{Ca}^{2+}/\text{Mg}^{2+}$ ratio^{29,30}. In addition, long lasting occlusion of this potentiation by LTP has recently been reported³¹. We found that the potentiation of m.e.p.s.cs remained ($35 \pm 5\%$ of control, $n = 4$) for at least 60 min in four of the five cells studied in elevated Ca^{2+} for this time. Furthermore, elevating Ca^{2+} considerably increased the magnitude of the potentiation of m.e.p.s.cs ($2.5 \text{ mM} = 53 \pm 13\%$, $n = 9$; $4 \text{ mM} = 60 \pm 8\%$, $n = 8$; $10 \text{ mM} = 151 \pm 31\%$, $n = 3$). NMDA-induced potentiation of m.e.p.s.cs

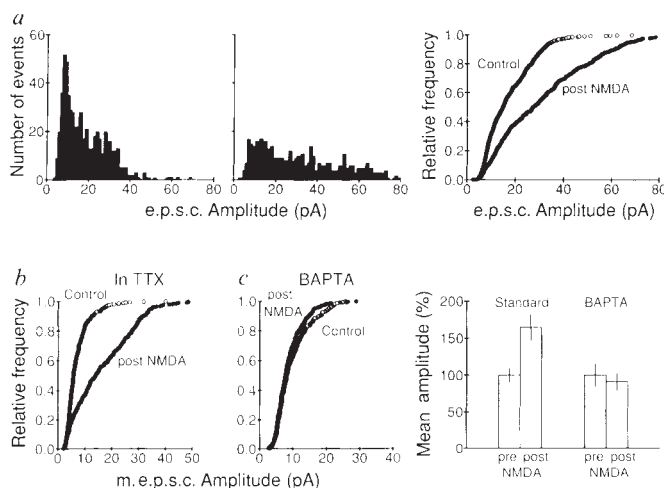


FIG. 6 Properties of the NMDA-induced increase in m.e.p.s.c. amplitude. *a*, Amplitude histograms of spontaneous e.p.s.cs recorded over a 5-min period immediately before (left), from 15 to 20 min after pressure application of NMDA (middle) in the absence of TTX, and cumulative plots of the same data (right). *b*, Cumulative plots from an experiment that was done entirely in TTX. NMDA ($10 \mu\text{M}$) was bath applied for 2 min. The superfusing medium contained 2.5 mM Ca^{2+} . The control curve comprises m.e.p.s.cs recorded over a 3-min period. The curve after NMDA comprises m.e.p.s.cs recorded 15 min after the application for a 3-min period. *c*, Example of an experiment in which the patch electrode contained 10 mM BAPTA (left). After pressure application of NMDA there was a small, but statistically insignificant, decrease in the amplitude of m.e.p.s.cs. Summary graph of the experiments which tested the effect of intracellular BAPTA on the NMDA-induced potentiation of m.e.p.s.cs (right). Methods as in Fig. 5. Standard results were obtained with the same internal solution as used in Fig. 5, whereas BAPTA results were obtained with an internal solution that contained 10 mM BAPTA . Potassium gluconate was reduced to maintain the same osmolarity and EGTA was omitted. The two types of experiments were alternated. The superfusing medium contained 4 mM Ca^{2+} .

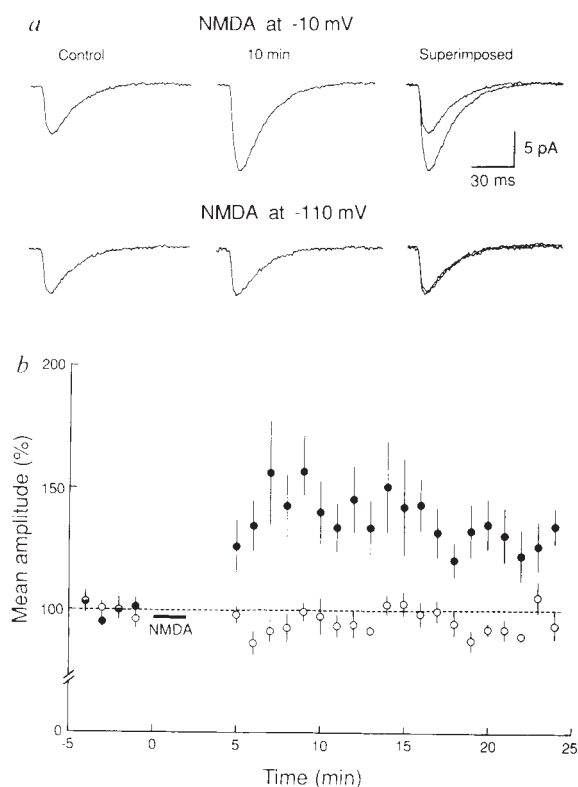


FIG. 7 Depolarization during NMDA application is required for potentiation of m.e.p.s.cs. *a* (top), Averaged records (40 traces) of m.e.p.s.cs obtained before (control), 10 min after bath application of NMDA ($10 \mu\text{M}$ for 2 min), and superimposed. During the application the cell was voltage clamped at -10 mV . *a* (bottom), The cell was voltage clamped at -110 mV during the application of NMDA (average of 35 traces). *b*, Summary of all of the experiments obtained with a superfusing medium containing 2.5 mM Ca^{2+} . ●, Results from cells voltage clamped at -10 mV during the application of NMDA ($n = 7$); ○, results from cells voltage clamped at -110 mV ($n = 5$). In all cases TTX was present throughout the experiment.

was found to share two other key properties of LTP: dependence on a rise in postsynaptic Ca^{2+} and on depolarization of the postsynaptic membrane. In contrast to the marked potentiation of m.e.p.s.cs by pressure application of NMDA ($60 \pm 8\%$, $n = 8$) that was recorded with electrodes containing 0.2 mM EGTA (which at this concentration weakly buffers Ca^{2+} and does not prevent LTP of evoked e.p.s.cs), when electrodes contained 10 mM of the Ca^{2+} chelator BAPTA, no potentiation was observed ($n = 4$) (Fig. 6c). Furthermore, voltage clamping the membrane potential at -110 mV during the bath application of NMDA also prevented the potentiation ($n = 5$) (Fig. 7a, bottom panel and *b*). The finding that chelating postsynaptic Ca^{2+} or preventing postsynaptic depolarization blocks the potentiation of m.e.p.s.cs induced by NMDA application not only suggests that this phenomenon is closely related to LTP, which is also blocked by these two manipulations, but also indicates that the effect of NMDA cannot be explained by a direct action on presynaptic structures. Furthermore, the observation that bath application of NMDA, especially in the presence of a high Ca^{2+} concentration, causes a reproducible and robust increase in the size of m.e.p.s.cs suggests that an LTP-like change occurs in a large number of synapses in the slice and therefore this procedure may provide a reliable method for measuring changes in biochemical processes related to LTP.

Discussion

The analysis of m.e.p.s.cs generally provides a simple and relatively direct method for detecting a change in postsynaptic transmitter sensitivity. Here we have applied this analysis to

synapses in the CNS and found that the induction of LTP causes an abrupt and lasting increase in the amplitude of m.e.p.s.cs. In addition, NMDA application causes a large increase in m.e.p.s.c. amplitude as well as in the size of the evoked e.p.s.c. This potentiation of m.e.p.s.cs shares many features with LTP. Our experiments suggest that a substantial portion of the potentiation of evoked e.p.s.cs can be accounted for by an increase in quantal size and by implication a change in postsynaptic transmitter sensitivity. We cannot exclude possible additional and separate presynaptic changes in the number of quanta released. Such changes would remain undetected by simply using m.e.p.s.c. amplitudes to determine the site responsible for a persistent change during LTP. Recently, on the basis of quantal analysis, evidence for such presynaptic changes has been presented (refs 12–15, but see ref. 16). But it is unclear whether the assumptions necessary to perform quantal analysis hold for CNS synapses^{17,18}.

In theory a change in quantal size could result from a change in the amount of transmitter released with each packet. For instance, at the vertebrate neuromuscular junction, depletion of the vesicular content of acetylcholine with the choline uptake blocker, hemicholinium, can reduce the size of quanta without altering postsynaptic sensitivity to acetylcholine³². But note that such a mechanism would still be interpreted as postsynaptic on the basis of analysis of fluctuations of evoked e.p.s.cs. Another class of presynaptic mechanisms could involve the cooperative linkage of release of multiple vesicles during LTP. Such a process could appear to be either pre- or postsynaptic depending on the details of the presynaptic interaction.

Our results, showing an increase in m.e.p.s.c. amplitude, indicate that either a postsynaptic modification contributes to LTP or alternatively a novel, unsuspected, presynaptic mechanism contributes to LTP. It will be important in future studies to distinguish between these two models. □

Received 29 July; accepted 7 November 1991.

- Bliss, T. V. P. & Lomo, T. *J. Physiol., Lond.* **232**, 331–356 (1973).
- Bliss, T. V. P. & Lynch, M. A. in *Long-Term Potentiation: From Biophysics to Behavior* (eds Landfield, P. W. & Deadwyler, S. A.) 3–72 (Liss, New York, 1988).
- Nicoll, R. A., Kauer, J. A. & Malenka, R. C. *Neuron* **1**, 97–103 (1988).
- Wigström, H., Gustafsson, B., Huang, Y.-Y. & Abraham, W. C. *Acta physiol. scand.* **126**, 317–319 (1986).
- Malinow, R. & Miller, J. P. *Nature* **320**, 529–530 (1986).
- Kelso, S. R., Ganong, A. H. & Brown, T. H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5326–5330 (1986).
- Sastry, B. R., Goh, J. W. & Auyeung, A. *Science* **232**, 988–990 (1986).
- Dolphin, A. C., Errington, M. L. & Bliss, T. V. P. *Nature* **297**, 496–498 (1982).
- Kauer, J. A., Malenka, R. C. & Nicoll, R. A. *Neuron* **1**, 911–917 (1988).
- Muller, D., Joly, M. & Lynch, G. *Science* **242**, 1694–1697 (1988).
- Davies, S. N., Lester, R. A. J., Reymann, K. G. & Collingridge, G. L. *Nature* **338**, 500–503 (1989).
- Voronin, L. L., Kuhnt, U. & Hess, G. *Neurophysiology* **22**, 341–347 (1990).
- Malinow, R. & Tsien, R. W. *Nature* **346**, 177–180 (1990).
- Bekkers, J. M. & Stevens, C. F. *Nature* **346**, 724–729 (1990).
- Malinow, R. *Science* **252**, 722–724 (1991).
- Foster, T. C. & McNaughton, B. L. *Hippocampus* **1**, 79–91 (1991).
- Edwards, F. *Nature* **350**, 271–272 (1991).
- Korn, H. & Faber, D. S. *Trends in Neurosci.* **14**, 439–445 (1991).
- Del Castillo, J. & Katz, B. *J. Physiol., Lond.* **124**, 560–573 (1954).

- Redman, S. *Physiol. Rev.* **70**, 165–198 (1990).
- Katz, B. *Nerve, Muscle and Synapse* (McGraw-Hill, New York, 1966).
- Bekkers, J. M., Richerson, G. B. & Stevens, C. F. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5359–5362 (1990).
- Van der Kloot, W. *Prog. Neurobiol.* **36**, 93–130 (1991).
- Barrett, E. F. & Stevens, C. F. *J. Physiol., Lond.* **227**, 691–708 (1972).
- Rahamimoff, R. & Yaari, Y. *J. Physiol., Lond.* **228**, 241–257 (1973).
- Fatt, P. & Katz, B. *J. Physiol., Lond.* **117**, 109–128 (1952).
- Collingridge, G. L., Kehrl, S. J. & McLennan, H. J. *J. Physiol., Lond.* **334**, 33–46 (1983).
- Kauer, J. A., Malenka, R. C. & Nicoll, R. A. *Nature* **334**, 250–252 (1988).
- Thibault, O. et al. *Brain Res.* **476**, 170–173 (1989).
- Malenka, R. C. *Neuron* **6**, 53–60 (1991).
- Asztely, F., Hanse, E., Wigström, H. & Gustafsson, B. *Brain Res.* **558**, 153–156 (1991).
- Elmqvist, D. & Quastel, D. M. J. *J. Physiol., Lond.* **177**, 463–482 (1965).
- Coleman, P. A. & Miller, R. F. *J. Neurophysiol.* **61**, 218–230 (1989).
- Blanton, M. G., Lo Turco, J. J. & Kriegstein, A. R. *J. Neurosci. Meth.* **30**, 203–210 (1989).

ACKNOWLEDGEMENTS. We thank colleagues at UCSF for their comments on the manuscript and D. Perkel and S. Hestrin for assistance with computer programming, especially for detecting and measuring spontaneous e.p.s.cs. This research was supported by a postdoctoral fellowship from the Uehara Memorial Foundation, Japan (to T. M.), the Swiss National Foundation for Scientific Research (to P.R.) and the NIH and NIMH (to R.A.N.). R.A.N. is a member of the Keck Center for Integrative Neuroscience and the Silvio Conti Center for Neuroscience Research.

LETTERS TO NATURE

Faint galaxies: evolution and cosmological curvature

Thomas J. Broadhurst*, Richard S. Ellis†
& Karl Glazebrook†

* Royal Observatory, Edinburgh EH9 3HJ, UK

† Physics Department, Durham University, Durham DH1 3LE, UK

RECENT observations of faint galaxies at near-infrared wavelengths^{1–3} reveal a surprisingly low surface density when compared to the excess of blue galaxies seen at optical wavelengths⁴. Attempts to determine the cosmological curvature from the asymptotic surface density of faint galaxies thus produce conflicting results^{3,5–7}. We propose to resolve this conflict with an evolutionary model in which galaxies merge at recent look-back times. The contrast between optical and infrared galaxy counts then follows from the very different lifetimes of stellar types contributing to emission in the galactic rest-frame. Together with evidence we present an increased star formation rate in galaxies at moderate redshift, the merging model can account for both the number–magnitude relations and available redshift distributions. A clear prediction is that there should be an absence of high-redshift galaxies in deep infrared-selected surveys. If correct, the model confirms earlier suspicions that galaxies cannot be used as reliable tracers of the geometry of the Universe.

The considerable excess of faint blue galaxies counted in the b_J passband ($\lambda_{\text{eff}} = 4,500 \text{ Å}$) over the numbers predicted from models involving no evolution⁴ (Fig. 1a) has attracted much

attention. Luminosity evolution in the galaxy population would affect the apparent magnitude, m , at which a given galaxy appears, but the integrated number to a given redshift depends sensitively on cosmic curvature or the deceleration parameter, q_0 , through volume effects, provided galaxy numbers are conserved with time. In principle, galaxy counts contain information on both aspects: the difficulty lies in disentangling their dependencies.

If it is assumed galaxies are drawn from a maximum accessible volume, either because they disappear from view at some distance, for example where the Lyman discontinuity of hydrogen is redshifted through the passband, or because there is an epoch earlier than which they did not exist, then the asymptotic surface density at faint limits (where the counts $dN(<m)/dm$ converge) can be compared with some estimate of the present number density to place constraints on q_0 irrespective of luminosity evolution. For the optical counts, the large asymptotic value⁵ yields a surprisingly low value of q_0 (ref. 6), in conflict with the standard inflationary model ($\Omega = 1$, $q_0 = 0.5$), unless a non-zero cosmological constant is considered⁷.

The situation has become more complicated with the arrival of the first galaxy counts using infrared arrays operating at $2 \mu\text{m}$ (the K-band)^{1–3}. Such a passband should be less affected by changes in the star formation rate because, in general, the contributing stars are older. Indeed, the K counts show a far less significant excess above the no-evolution prediction (Fig. 1b) and, although the asymptotic surface density is not yet well determined, the counts seem to favour a high value of q_0 (ref. 3).

Here we show that a straightforward model for both the optical b_J and infrared K counts can, in principle, be consistent with

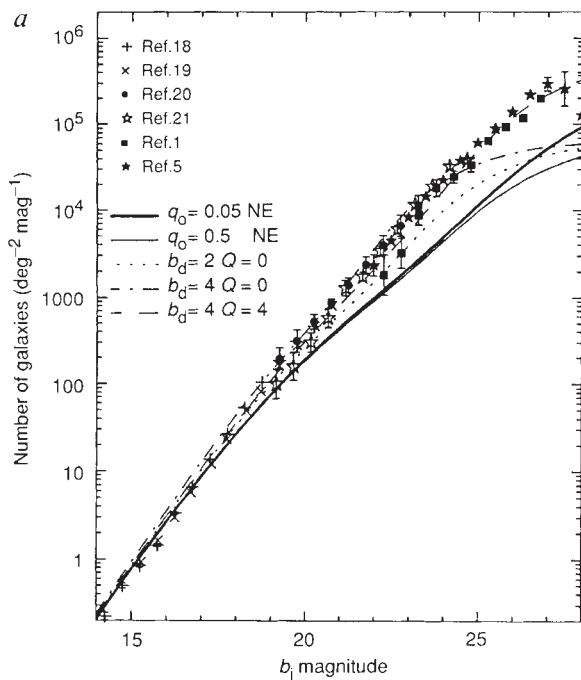
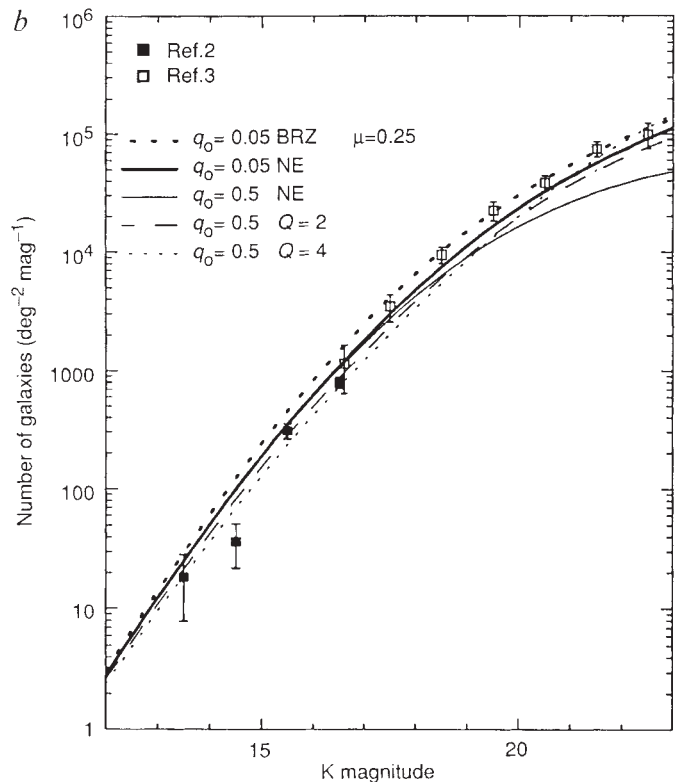


FIG. 1 Differential galaxy counts in (a) the optical b_J and (b) the infrared K passbands from various sources^{1-3,5,18-21}. Solid lines refer to no evolution, (NE) predictions for $q_0=0.05$ and 0.5 based on present-day properties. Other models incorporate various rates of galaxy merging (Q) and a rise in star formation rate (b_d). Star formation is modelled according to a standard



initial mass function, and evolutionary corrections in the infrared are based on Bruzual's evolutionary code¹⁰ (BRZ).

any reasonable value of q_0 , including that preferred in the inflationary picture ($q_0=0.5$). The starting point for our discussion is to abandon the standard assumption that galaxy numbers are conserved with cosmic time. Merging in the galaxy populations will clearly lead to an increase in the galaxy density with look-back time. Furthermore, if the associated star formation rate in the merging population increases similarly, we show this has different consequences for the b_J -selected and K-selected samples. Our model is an extension of one proposed to explain the surprising results of two recent b_J -limited spectroscopic surveys^{8,9}. These surveys provide valuable new data, from emission-line measurements, on the rate of star formation for several hundred faint galaxies.

The main difficulty for the evolutionary models is the surprisingly shallow depth of the faint b_J -selected redshift surveys. Because the counts are much in excess of the no-evolution prediction at these limits, then the blue luminosity density must increase with look-back time⁷. In addition, an increasing proportion of the faint spectra reveal enhanced star formation, as quantified by the rest-frame equivalent width, W_λ , of the commonly-detected [O II] emission line at 3,727 Å. Completion of a survey in the intermediate magnitude range $17 < b_J < 20.5$ (T.J.B. and R.S.E., manuscript in preparation) now provides sufficient data to demonstrate that the steep slope of the galaxy counts arises directly from these star-forming galaxies (Fig. 2).

To demonstrate how merging and a declining star formation rate can resolve the optical/infrared paradox, we have constructed some simple models, which have the useful feature of not depending critically on particular details of the stellar evolutionary models (for example ref. 10). We assume merging occurs in a self-similar way for all objects subject to the constraint that the total mass of associated material is conserved. For simplicity, we assume all galaxies merge but the proportions of different types, their respective K-corrections and luminosity function shapes are maintained. A characteristic mass for the self-similar galaxy mass function, say M^* clearly decreases with increasing

look-back time, δt , at the expense of an increased number density, Φ^* . If $\Phi^*(\delta t) = f(\delta t)\Phi_0^*$ then $M^*(\delta t) = f(\delta t)^{-1}M_0^*$, where the subscript '0' refers to present-day values. We adopt an exponential form for $f(\delta t)$:

$$f(\delta t) = \exp(Q\delta t/\beta t_0) = \exp(-Q/\beta((1+z)^{-\beta} - 1)) \\ \approx 1 + Qz \quad (z < 1)$$

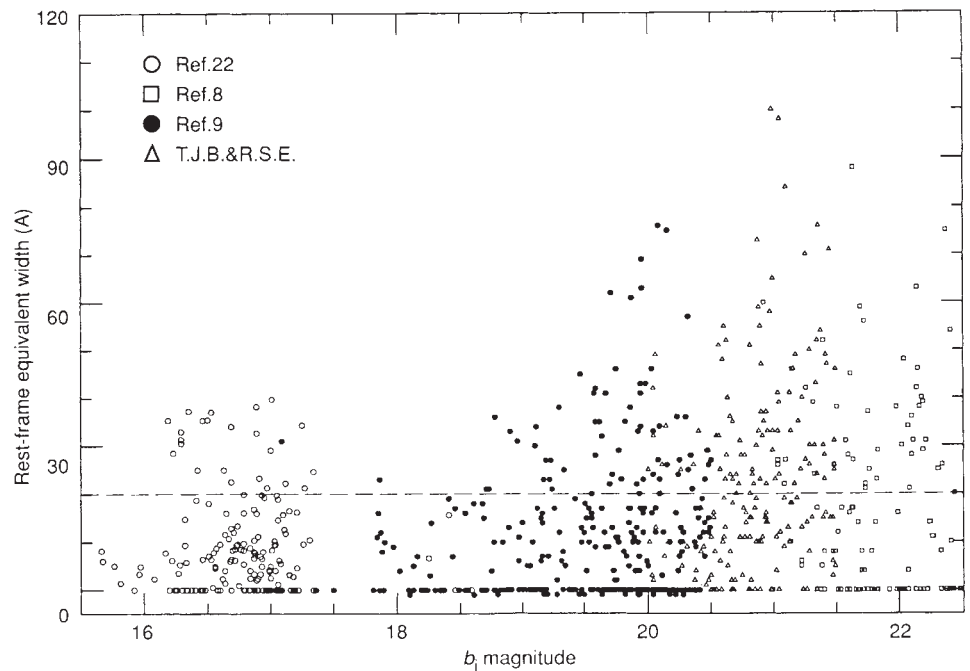
where Q is a parameter defining the merger rate and $\beta = 1 + (2q_0)^{0.6}/2$ (see ref. 11). This formalism is preferable to that where merging proceeds exponentially with $(1+z)$ (ref. 12) because the rate does not become unreasonably high at very early times. A model with $Q=4$, for example, implies that a present-day galaxy is the result of a merger between ~ 4 – 6 sub-units which existed at $z=1$, depending on q_0 . Although more galaxies are present per unit volume at larger distances, each is less luminous and the effects on the counts cancel to first order. Merging does, however, have an important effect on the redshift distributions, reducing the average redshift of a faint survey as required to explain the observations (Fig. 3a).

The luminosity of a distant galaxy also depends on the history of its star formation rate, $\Psi(\delta t)$. Optical counts will be affected mainly by hot short-lived stars, but those in the K-band will be less affected, as the dominant stellar types are longer-lived. From Fig. 2 we now have good evidence that Ψ was, on average, greater in the past. Under the reasonable assumption that Ψ is proportional to the gas available, then the mean star formation rate per unit mass will evolve exponentially as:

$$\Psi(\delta t) = \Psi_0 \exp(b_d \delta t / \beta t_0)$$

where b_d is a measure of the average rate of decline. The increase in blue luminosity is closely proportional to Ψ at modest redshifts because the ultraviolet radiation enters the optical band. This formalism is similar to the μ -models of Bruzual¹⁰ in

FIG. 2 The rapid increase with apparent magnitude in the number of star-forming galaxies as quantified by rest-frame equivalent widths of the $[O II]$ emission line at $3,727 \text{ \AA}$ from recent redshift surveys (refs 8, 9, 22, and T.J.B. and R.S.E., in preparation). Dotted line, the upper limit ($\sim 20 \text{ \AA}$) seen in local spirals. The count slope, $\gamma = d \log N / d m$ over the range $18 < b_j < 21$, when normalized for the different survey sampling rates, is roughly Euclidean ($\gamma = 0.6$) for galaxies with lines stronger than this value, whereas the remainder share $\gamma \approx 0.35$, the predicted no-evolution value.



which $b_d = -3t_0(\text{Gyr}) \ln(1 - \mu)/2$. For $t_0 = 13 \text{ Gyr}$ ($q_0 = 0.5$), a galaxy evolving with $b = 4$, $\mu \approx 0.25$, for example, could be 10 Gyr old and have a 10% gas content today.

The optical counts are strongly affected if $b_d > 0$, but without merging there would be an unobserved high redshift population in the faintest surveys. A satisfactory feature is our ability to constrain b_d from the counts and Q , largely independently, from the redshift data (Figs 1a and 3a). A combination of $b_d > 0$, $Q > 0$ is clearly required and, under the simple assumptions of the model, $Q = 4$ and $b_d = 4$ provide a good fit. We note that similar rates of evolution for b_d and Q may indicate that the increased star formation is, indeed, associated in some way with the merging process.

Of course, an increased star formation rate has a much smaller effect at K. Although the detailed effects remain unclear because the post-main sequence stellar contributions are uncertain, we can estimate what will happen in the case where Ψ evolves according to the optical fit, $\mu = 0.25$, in the context of Bruzual's model¹⁰. At $z \approx 1$, the K evolutionary corrections range from 0.07–0.45 mag in K, depending on q_0 , compared with 1.7 mag in b_j . Under these assumptions $N(K)$ is largely insensitive to both merging and past star formation (Fig. 3b). Sub-unit galaxies lose their visibility very rapidly, whereas in b_j the increased star formation more than compensates. Specifically, we would expect the mean redshift of a K-selected sample to be less than a no-evolution prediction (Fig. 3b). Earlier attempts to reconcile the faint galaxy data do not make this prediction^{3,8}, because the evolution is superimposed upon an unchanging galaxy mass function. This would require extremely blue $B - K$ colours in order to be consistent with the data. For example, a model that adds a dwarf galaxy population at earlier times would require those galaxies to have $B - K \leq 2$ (whereas $B - K \approx 4$ is observed (ref. 3)), so that the faint B excess at $B > 22$ does not also appear in the K counts in the relatively well-established regime $K < 20$.

Our model applies to the general field population and does not necessarily conflict with the presence of luminous rich cluster members at early times, as hierarchical evolution of the kind described would proceed more rapidly in dense environments¹³. We admit, however, that many details are missing, including independent measures of the merging rate today and at recent times, and the question of the stability of disk systems under such encounters¹⁴. As the K counts improve, a more detailed

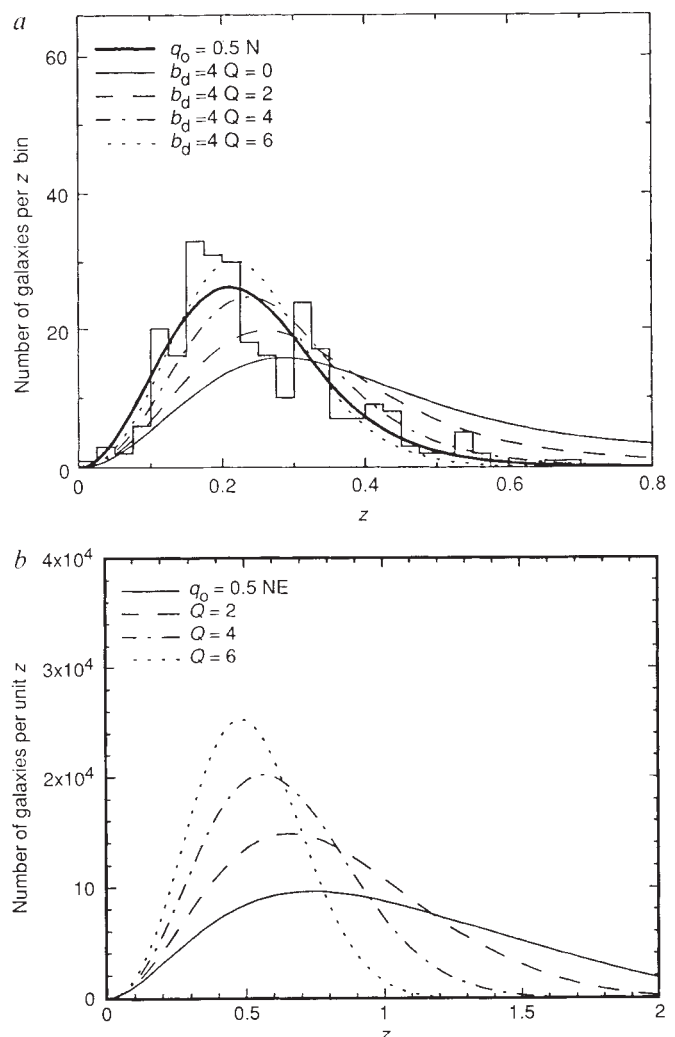


FIG. 3 Redshift distributions for (a) the combined b_j -selected sample of refs 8 and 9 limited at 22.5 with available data shown as a histogram, (b) a hypothetical survey limited at $K = 19.5$. Curves are as in Fig. 1.

treatment of the infrared luminosity evolution incorporating the post-main-sequence stages will be required.

Merging is a natural feature of conventional models for the growth of structure by gravity, so we expect a decline in the amplitude of the galaxy correlation function with look-back time. Non-conservation of galaxy numbers would not change this prediction, as the correlation function is normalized to the mean density at any epoch. On small scales, merging might increase the number of close pairs but the effect would depend on the merger timescale (which could be two orders of magnitude less than the observed look-back time) and whether galaxies brighten in luminosity before or during the event. Recent estimates of the projected angular correlations to $b_J = 26$ suggest an anomalously low amplitude of clustering, and claims are made for a new weakly clustered population of additional sources¹⁵. The small field areas measured at these limits do, however, involve making a large and uncertain boundary correc-

tion which is comparable to the claimed discrepancy in amplitude. Furthermore, earlier more extensive correlation studies to $b_J = 24$, which are not subject to such large corrections, and where the count excess is already an order of magnitude over no-evolution predictions (Fig. 1a), do not require a new population of sources^{16,17}. Given that the observational discrepancy is currently restricted to the uncertain CCD data, we suggest that more extensive data are required before detailed comparisons can be made.

Our purpose here has been to demonstrate the striking dynamical evolution can have on the faint galaxy distributions and how, with the recent spectroscopic evidence for evolution in the star formation rate provided by the redshift surveys, it can reconcile the apparent contradictory results of the optical and infrared counts. A consequence of this picture is the obvious difficulty of using galaxy counts to determine the curvature of space. □

Received 17 July; accepted 30 October 1991.

- Lilly, S. J., Cowie, L. L. & Gardner, J. *Astrophys. J.* **369**, 79–105 (1991).
- Glazebrook, K. thesis, Univ. of Edinburgh (1991).
- Cowie, L. L., in *Observational Tests of Inflation* (ed. Shanks, T.) (Kluwer, Dordrecht) 25–30 (1991).
- Ellis, R. S. in *Hubble Centennial Symp.* (ed. Kron, R.) *ASP Conf. Series*, 248–264 (1990).
- Tyson, A. J. *Astr. J.* **96**, 1–23 (1988).
- Yoshi, Y. & Takahara, F. *Astrophys. J.* **326**, 1–18 (1988).
- Cowie, L. L. & Lilly, S. J. in *Hubble Centennial Symp.* (ed. Kron, R.) *ASP Conf. Series* 212–220 (1990).
- Broadhurst, T. J., Ellis, R. S. & Shanks, T. *Mon. Not. R. astr. Soc.* **235**, 827–856 (1988).
- Colless, M. M., Ellis, R. S., Taylor, K. & Hook, R. H. *Mon. Not. R. astr. Soc.* **244**, 408–423 (1990).
- Bruzual, G. *Astrophys. J.* **273**, 105–127 (1983).
- Peacock, J. A. in *Astrophysical Jets and their Engines* (ed. Kundt, W.) 171 (Reidel, Dordrecht, 1987).
- Rocca-Volmerange, B. & Guiderdoni, B. *Mon. Not. R. astr. Soc.* **247**, 166–172 (1990).
- Bower, R. G. *Mon. Not. R. astr. Soc.* **248**, 332 (1991).

- Ostriker, J. P. in *Hubble Centennial Symp.* (ed. Kron, R.) *ASP Conf. Series* 1–20 (1990).
- Efstathiou, G., Bernstein, G., Katz, N., Tyson, J. A. & Guhathakurta, P. *Astrophys. J.* **380**, L47–50 (1991).
- Koo, D. C. & Szalay, A. *Astrophys. J.* **282**, 390–397 (1984).
- Stevenson, P. R. F., Shanks, T., Fong, R. & MacGillivray, H. T. *Mon. Not. R. astr. Soc.* **213**, 953–969 (1985).
- Collins, C. A. & Nichol, R. *Mon. Not. R. astr. Soc.* (in the press).
- Maddox, S. J., Sutherland, W. J., Efstathiou, G., Loveday, J. & Petesson, B. A. *Mon. Not. R. astr. Soc.* **247**, 1 4P (1990).
- Jones, L. R., Fong, R., Shanks, T., Ellis, R. S. & Peterson, B. A. *Mon. Not. R. astr. Soc.* **249**, 481–497 (1991).
- Metcalfe, N., Shanks, T. & Fong, R. *Mon. Not. R. astr. Soc.* **249**, 498–522 (1991).
- Peterson, B. A. *et al.* *Mon. Not. R. astr. Soc.* **221**, 233–255 (1986).

ACKNOWLEDGEMENTS. We acknowledge discussions with R. Bower, C. Frenk, N. Metcalfe, G. Oemler, J. Peacock, I. Tyson, N. Weir and S. White. This work is supported by SERC.

Dynamical development of a ring-like structure in the supernova 1987A nebula

Lifan Wang*† & Paolo A. Mazzali*

* European Southern Observatory, Karl-Schwarzschild-Strasse 2, W-8046 Garching bei München, Germany
† Center for Astrophysics, The University of Science and Technology of China, Hefei, Anhui, China

CONTINUING observations of the nebula around SN1987A, which is about 3 arcseconds across, show it to be surprisingly regular. The brightest part of it is a highly axisymmetric structure, perhaps a simple ring^{1–4}. This structure, which may be related to similar structures in planetary nebulae and around Wolf–Rayet stars, poses a challenge for current nebular formation theories. Here we discuss the formation of an axisymmetric nebula in terms of the interaction between the slow wind ejected by the progenitor of SN1987A while it was a red supergiant (RSG) and the subsequent fast wind produced by the same star after it had evolved into a blue supergiant (BSG). In our model, an initial small asymmetry in the RSG wind, due to rotation of the red giant, is amplified by the action of the fast BSG wind. Structures that are significantly axisymmetric, and perhaps even ring-like, can be produced. We suggest that such phenomena should be common in the formation of nebulae, as evidenced by the fact that only a small fraction of the observed circumstellar nebulae appear to be spherically symmetric.

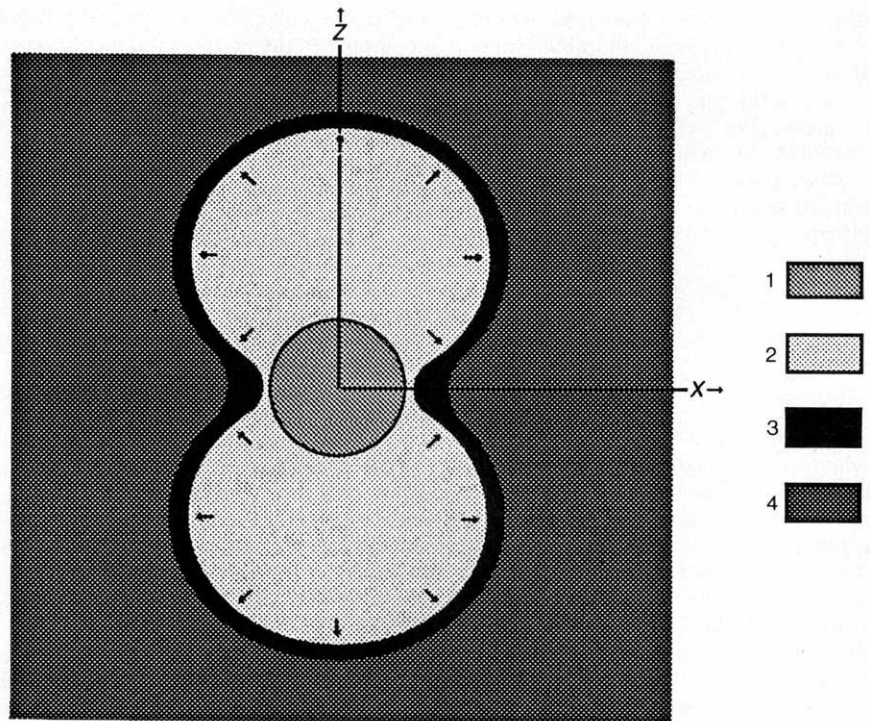
Circumstellar material of size $\sim 2''$ around SN1987A was first detected spectroscopically by the International Ultraviolet Explorer⁵, and then in the optical⁶. Imaging showed a nebula elongated in the east–west direction, which was then interpreted as light echoes from an inclined cylinder¹. Later observation

with the European Southern Observatory's New Technology Telescope (NTT) showed a well-structured nebula, consisting basically of two loops, the other one of which is about ten times fainter than the inner one. Because their size did not change appreciably with time, it was concluded that the loops represent the actual physical structure of a circumstellar nebula². An image taken with the Hubble Space Telescope (HST) showed the inner loop more clearly³. It is now believed that the nebula has a high degree of axial symmetry, possibly being a pure ring in three dimensions.

If Sk-69 202, the progenitor of SN1987A, exploded as a BSG after spending some time as a RSG, as is commonly thought⁷, the dynamical origin of the SN1987A circumstellar structure is analogous to that of circumstellar bubbles⁸. As a RSG, the star ejected mass at a high rate ($\sim 10^{-5} M_{\odot} \text{ yr}^{-1}$), but with a low velocity ($\sim 10 \text{ km s}^{-1}$). Later, in the BSG stage, the mass loss rate is smaller ($\sim 2 \times 10^{-6} M_{\odot} \text{ yr}^{-1}$), but a higher velocity ($\sim 500 \text{ km s}^{-1}$) is reached. The BSG wind thus quickly catches up with the RSG wind. The successive interaction between the two winds gives rise to a four-zone structure^{8,9}. At increasing radii, the following regions are encountered: (1) the hypersonic BSG wind; (2) a hot ($T \approx 10^7 \text{ K}$) and thus almost isobaric region of shocked fast wind, caused by the sudden deceleration of the fast wind; (3) a thin, dense, cold and also shocked shell containing most of the swept-up slow wind; (4) the still-unshocked portion of the slow wind. The observable circumstellar structure is thought to arise from region 3 (ref. 9). A similar mechanism is believed to be responsible for the formation of planetary nebulae (PNe), and the morphological similarity between the SN1987A nebula and PNe, noted by Wampler *et al.*², is thus not surprising.

As the SN1987A nebula and most other nebular structures are not spherically symmetric, but show signs of axial symmetry, it is interesting to investigate how axially symmetric nebular shapes can develop in the context of this interaction model. Here we assume that the axial symmetry originates in an asymmetry of the slow RSG wind.

FIG. 1 Schematic picture of the interaction of the fast wind ejected during the BSG stage with the slower wind ejected during the RSG stage. The formation of a hot, shocked fast wind region (2), due to a reverse shock, causes the force (arrows) on the shocked slow wind region (3) to be normal to the contact discontinuity between the two shocked regions.



Following Kahn and West¹⁰, we describe the dependence of the slow wind density (ρ_R) on stellar latitude (θ) by

$$\rho_R(r, \theta) = \frac{\dot{M}_R(\theta)}{4\pi r^2 v_R} \quad (1)$$

where

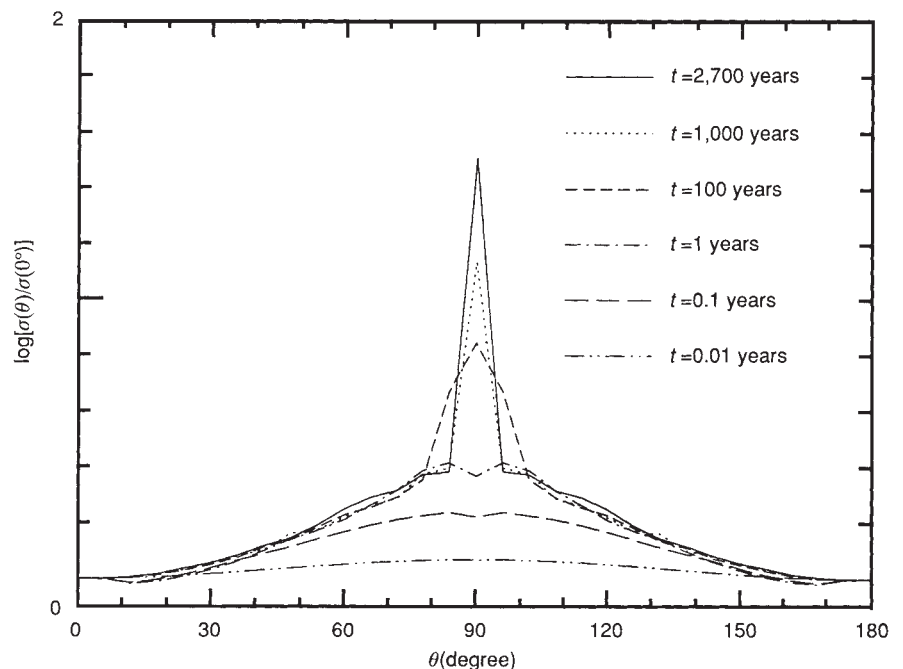
$$\dot{M}_R(\theta) = \dot{M}_R(\pi/2) \cdot (1 - A \cos^2(\theta)) \quad (2)$$

is the mass loss rate of the progenitor star during the RSG stage, r is the radius to the origin and v_R is the velocity of the slow wind. A is thus a measure of the deviation from spherical symmetry. As the star evolves from RSG to BSG, which is a rapid process, the wind properties take the typical values $v_B = 500 \text{ km s}^{-1}$ and $\dot{M}_B = 2 \times 10^{-6} M_\odot \text{ yr}^{-1}$.

As the fast wind runs into the slow wind, a four-zone structure as described above develops. The overall picture of the interac-

tion is illustrated in Fig. 1. In the presence of an asymmetry, one interesting feature arises, which determines the successive development of the nebular shape. The temperature of region 2 is typically $\sim 10^7 \text{ K}$, so that the local sound speed is much higher than the velocity of the shocked slow wind. The shocked fast wind is thus almost isobaric when interacting with the shocked slow wind. The force pushing region 3 is then normal to the contact discontinuity between regions 2 and 3. As long as the reverse shock can develop over the entire 4π solid angle, any asymmetry of the fast wind is then not important in shaping the nebula. The fast wind does not interact directly with the slow wind, but does so through an isobaric region. The force normal to the shock surface amplifies any deviations from spherical symmetry of the slow wind while conserving angular momentum. With the RSG wind density distribution given by equation (1), the fast wind pushes more effectively in the polar

FIG. 2 The surface density of the shocked slow wind at different epochs after the onset of the fast wind. The values have been normalized to the values in the polar direction. The density enhancement on the equatorial plane due to the small initial asymmetry is responsible for the SN1987A ring nebula. The initial conditions adopted for the calculations are: for the fast wind¹⁵, $\dot{M}_B = 1.7 \times 10^{-6} M_\odot \text{ yr}^{-1}$, $v_B = 550 \text{ km s}^{-1}$; for the slow wind¹⁵, $\dot{M}_R = 1 \times 10^{-5} M_\odot \text{ yr}^{-1}$, $v_R = 5 \text{ km s}^{-1}$, and $A = 0.2$.



direction, giving rise to the typical hourglass structure for region 3. The pressure of the hot bubble (region 2) compresses the two sides of the shocked RSG wind onto the equatorial plane, so that they will collide. A high-density cusp subsequently develops. We believe that this cusp is responsible for the SN1987A ring nebula.

If we assume the shocked slow wind to be infinitely thin, transverse pressure gradients within this region can be ignored. After a simple derivation, the equations of motion are found to be

$$\sigma \frac{d\mathbf{u}}{dt} = P\mathbf{n} - \rho_R(u_n - v_{Rn})(\mathbf{u} - \mathbf{v}_R)$$

and

$$\frac{d \ln(\sigma S)}{dt} = \frac{\rho_R(u_n - v_{Rn})}{\sigma} - \frac{\partial \theta}{\partial \theta} \quad (4)$$

where $S = r[r^2 + (\partial r / \partial \theta)^2]^{1/2} \sin \theta$ is the area on the shocked slow wind surface subtended by unit solid angle, σ is the surface mass density of the shocked slow wind, $\mathbf{u}(\mathbf{r})$ is the velocity of a particular element on the shocked surface, P is the pressure inside the shocked fast wind, and $\mathbf{n}$ is the unit vector normal to the shock surface. The subscript n stands for the normal component of a vector. A further equation can be derived by equating the energy input from region 1 into region 2 to the work done by region 2 against region 3. As the thermal energy of region 2 is much larger than its kinetic energy, the total energy of this region is given by

$$\frac{d}{dt}(PV) = \frac{2}{3} \left(\dot{E}_0 - 4\pi \int_0^{\pi/2} P \cdot u_n S d\theta \right) \quad (5)$$

where V is the volume enclosed by the shocked slow wind and $\dot{E}_0$ is the energy input from the fast wind to region 2. The model therefore depends only on four parameters: $\dot{E}_0$, $\dot{M}_R(\frac{\pi}{2})$, v_R and A .

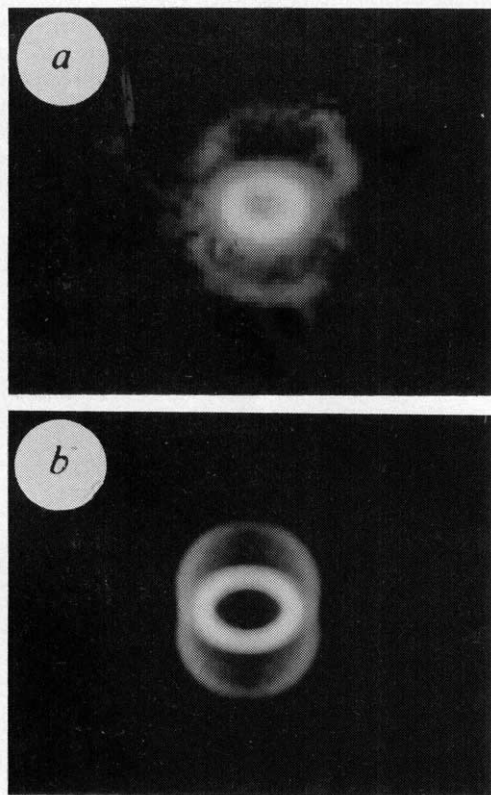


FIG. 3 *a*, NTT image of the supernova in the light of the narrow band [O III] 501-nm filter. *b*, Simulated image of the ring nebula. The angle between the observer and the symmetry axis is 43° , as measured in the HST image³. The outer loop is due to the limb-brightening of the hourglass-shaped shell.

Solving these equations, we modelled the evolution of the interaction region after the fast wind starts to push the slow wind. The initial conditions adopted for the calculations are given in Fig. 2. The collision between the two sides of the shocked slow wind on the equatorial plane was assumed to be completely inelastic and no mass outflow was allowed from the equatorial plane.

The surface mass density of the shocked slow wind for different epochs after the onset of the fast wind is shown in Fig. 2. The mass on the equatorial plane increases rapidly over the first few years, mostly because of the forced motion of the matter towards that plane. The outer lobe attains a steady configuration in ~ 500 years while the ratio of equatorial to polar mass is still increasing. The time required to develop a highly axially symmetric structure, with $\sigma(\theta \leq 80^\circ) / \sigma(90^\circ) \leq 1/5$, is $\sim 1,000$ years, much less than the time the progenitor of SN1987A spent as a BSG.

Because recent observations have shown that the SN1987A nebula is still at a temperature higher than 20,000 K, and the light travel effect is not important for the late morphology of the nebula¹¹, the emission measure is obtained by simply computing the square of the particle number density along each line of sight. The dynamical age of the ring nebula is constrained by a recent NTT-EMMI (ESO Multimode Instrument) echelle spectrum to be 27,000 years (L. Wang *et al.*, manuscript in preparation), so our simulation ends at $t = 27,000$ years after the onset of the BSG wind. The simulated structure, shown in Fig. 3*b*, is observed as two concentric loops, looking remarkably similar to the NTT image of SN1987A shown in Fig. 3*a*. As the thin shock approximation breaks down on the equatorial plane, the calculated equatorial velocity and hence the size of the ring are not accurate, but the mass of the ring is correctly calculated. The velocity in the polar direction is found to be 56 km s^{-1} at $t = 27,000$ years.

A recently reported model¹², based on a similar physical mechanism, adopted an initial ratio between the equatorial and the polar mass loss rates of the slow wind of 5:1 ($A = 0.8$ in our model), producing an equator-to-pole column density ratio of 200:1 at $t = 8,000$ years. Such a large value of A is hard to justify for two reasons: first, the observed intensity ratio between the inner and outer loop is only ~ 10 – 20 , much lower than the value they would obtain; second, as the age of the nebula is much more than 8,000 years, this ratio will increase further if calculated for the correct age.

It has been suggested that the reason for the seed asymmetry could be the presence of a companion star^{13,14}. Although this may be true in the case of SN1987A, the small A value required in our model indicates that the possible role of rotation cannot be firmly excluded. A comparison with PNe lends support to this. Most PNe are very well structured, but few of them are spherically symmetric. Although many PNe have binary nuclei, it is hard to believe that all axially symmetric nebulae are due to binaries. It is still interesting to investigate the role of rotation on the mass loss rate of the RSG wind.

We conclude that the morphology of the SN1987A ring nebula can be nicely reproduced by a model involving the interaction of slow and fast winds, if the density of the slow RSG wind is assumed to be slightly enhanced on the equatorial plane. The two-wind interaction amplifies this asymmetry. A similar mechanism is probably fairly common in nature, being responsible for the various morphologies of planetary nebulae and the ring nebulae surrounding Wolf-Rayet stars. $\square$

Received 5 June; accepted 15 October 1991.

1. Crotts, A. P. S., Kunkel, W. E. & McCarthy, P. J. *Astrophys. J.* **347**, L61–64 (1989).
2. Wampler, E. J. *et al. Astrophys. J.* **362**, L13–16 (1990).
3. Jakobson, P. *et al. Astrophys. J.* **369**, L83–86 (1991).
4. Crotts, A. P. S. & Heathcote, S. R. *Nature* **350**, 683–685 (1991).
5. Fransson, C. *et al. Astrophys. J.* **336**, 429–441 (1989).
6. Wampler, E. J. & Richichi, A. *Astr. Astrophys.* **217**, L21–24 (1989).

7. Maeder, A. in *Proc. ESO Workshop on SN1987A* (ed. Danziger, I. J.) 251–270 (Reidel, Dordrecht, 1987).
8. Castor, J., McCray, R. & Weaver, R. *Astrophys. J.* **200**, L107–110 (1975).
9. Chevalier, R. A. *Nature* **332**, 514–516 (1988).
10. Kahn, F. D. & West, K. A. *Mon. Not. R. astr. Soc.* **212**, 837–850 (1985).
11. Wang, L. *Astr. Astrophys.* **246**, L69–72 (1991).
12. Luo, D. & McCray, R. *Astrophys. J.* (in the press).
13. Chevalier, R. A. & Soker, N. *Astrophys. J.* **341**, 867–882 (1989).
14. Podsiadlowski, P. *Nature* **350**, 654 (1991).
15. Fransson, C. & Chevalier, R. A. *Nature* **328**, 44–45 (1988).

ACKNOWLEDGEMENTS. We thank D. Baade and L. B. Lucy for discussions.

Detection of a new white-dwarf binary system in the extreme ultraviolet using the Rosat Wide Field Camera

B. A. Cooke*, **M. A. Barstow***, **E. R. Breeveld†**,
P. A. Charles‡, **C. V. Goodall§**, **J. A. Gourlay||**,
B. J. M. Hassall‡, **H. E. Huckle†**, **B. J. Kent||**,
J. P. Pye*, **A. G. Richards||**, **M. R. Sims***,
T. J. Sumner¶, **M. G. Watson***, **A. Wells***,
R. Willingale* & **J. S. Wright||**

* Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK

† Mullard Space Science Laboratory of University College, London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK

‡ Royal Greenwich Observatory, Madingley Road, Cambridge CB3 0EZ, UK

§ School of Physics and Space Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

|| Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, UK

¶ Department of Physics, Imperial College of Science, Technology and Medicine, Prince Consort Road, London SW7 2BZ, UK

DETAILED understanding of the evolution of main-sequence stars into white dwarfs depends on knowledge of the chemical composition of white-dwarf atmospheres, but the hottest white dwarfs emit much of their radiation in the extreme ultraviolet (EUV), a part of the spectrum that has been generally inaccessible to astronomers. We report here the discovery of a new white dwarf as a bright EUV source found during the first observations using the Wide Field Camera¹ on the Rosat satellite. Within the 1-arcmin error box of the EUV source is an unusually blue star, which we show to be a binary system consisting of a DA white dwarf and a cool companion in the classification range dM2–dM5. There are some similarities between this new object and the binary system Feige 24 (ref. 2), but spectral differences in the EUV emission can be attributed to significantly different atmospheric compositions for the two white dwarfs.

The Rosat satellite, launched 1 June 1990, contains two telescopes. One, the Wide Field Camera (WFC)¹ was built by a consortium of the University of Leicester, University of Birmingham, the Mullard Space Science Laboratory of University College, London, the Imperial College of Science, Technology and Medicine and the Rutherford Appleton Laboratory. The WFC has a field of view 5° in diameter and the observations described here were made in a total energy range of 62–190 eV. The other, the German X-ray Telescope (XRT)³ covers the energy range 100–2,400 eV with a co-aligned field of 2° diameter.

On 17 June 1990, when the XRT was making a calibration observation, the WFC, in parallel, was making 'first-light' observations to test the detectors and science filters. In the field we found a new EUV source which at ~1 WFC count s⁻¹ has subsequently been found to have an intensity among the seven brightest of all the EUV sources seen in the WFC all-sky survey. It has also turned out to be an unusual example of a white-dwarf binary system. More data were obtained when 11 observations,

approximately equally spaced in time, made through the WFC S1a filter (90–190 eV or 65–138 Å) produced an exposure of 14,500 s from a total elapsed time of 100,000 s. An additional observation of 300 s was made through the WFC S2b filter (62–111 eV or 112–200 Å). The resulting EUV light curve in the WFC S1a filter band shows no evidence for variability on timescales ranging from 100s to 12 h.

It is possible to position the source within a 90% confidence contour radius ~1 arcmin centred at (equinox J2000) coordinates (16 h 29 min 29 s, +78° 04' 31"). Six objects lie in or near the error circle, but one is brighter than the others by more than 5 magnitudes. These six objects are labelled 1–6 on the colour-magnitude diagram in Fig. 1. A combination of the stellar luminosity function we measure for objects within 2° of RE1629+781 (RE represents Rosat EUV) and the EUV error circle yields a formal probability of ~93% that object 1 (Poss O plate magnitude, $m_O = \sim 12.7$) is the counterpart, compared with much lower probabilities for objects 2 to 6, all of which are fainter than $m_O \approx 18$ mag. Figure 1 shows that object 1, which has (J2000) coordinates (16 h 29 min 11.1 s, +78° 04' 41.2") is outstandingly brighter and bluer with $m_O - m_E = 0.05$ ($m_O - m_E \approx 2(m_B - m_V)$) than any of the other objects in, or near, the error box. The vast majority of the stars in this region of the sky have typically $m_O - m_E \approx 1$. This places object 1 on the colour-magnitude diagram close to the position of the DA white dwarf HZ43, the brightest EUV source in the sky⁴.

Low-resolution (5 Å) spectroscopy of object 1 obtained with the William Herschel Telescope (WHT) using the intermediate-dispersion spectroscopic imaging system (ISIS) shows (Fig. 2a) that the blue part of the spectrum contains the characteristic, broad hydrogen absorption lines of a DA white dwarf. But the red spectrum (Fig. 2b) shows clear evidence of TiO or Na absorption at 5,900 Å and a pseudo-emission hump of radiation surrounding H α caused by TiO absorption at 6,200 Å and 6,900 Å, features suggesting that the white dwarf has a cool companion of spectral class M. The narrow emission core to H α (possibly present in H β as well) and the probable Ca K emission can be explained if EUV radiation from the white dwarf is being reprocessed on the surface of this cool companion. Comparison of these spectra with those of Liebert and Margon⁵

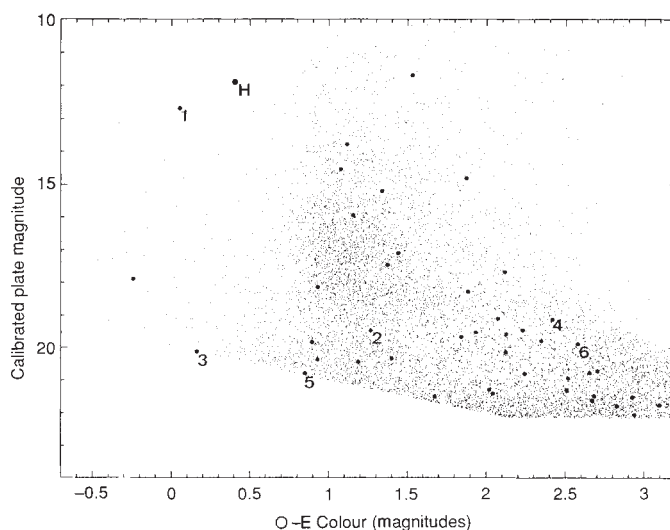


FIG. 1 Characteristics of the optical field associated with RE1629+781. Colour-magnitude diagram obtained from using the Automatic Plate Measuring Machine at Cambridge to measure the POSS O and E plates. The magnitudes are calibrated statistically using standard machine algorithms¹⁴ and have an internal consistency of ~0.3 mag. The small dots represent all (15,735) objects in a (2° × 2°) square centred on RE1629+781 and the superposed large dots the stars in a (6' × 6') square. The six stars within, or close to, the 90% confidence radius of the EUV position are numbered. The probable optical counterpart (object 1) lies close to HZ43 (labelled H) in the diagram.

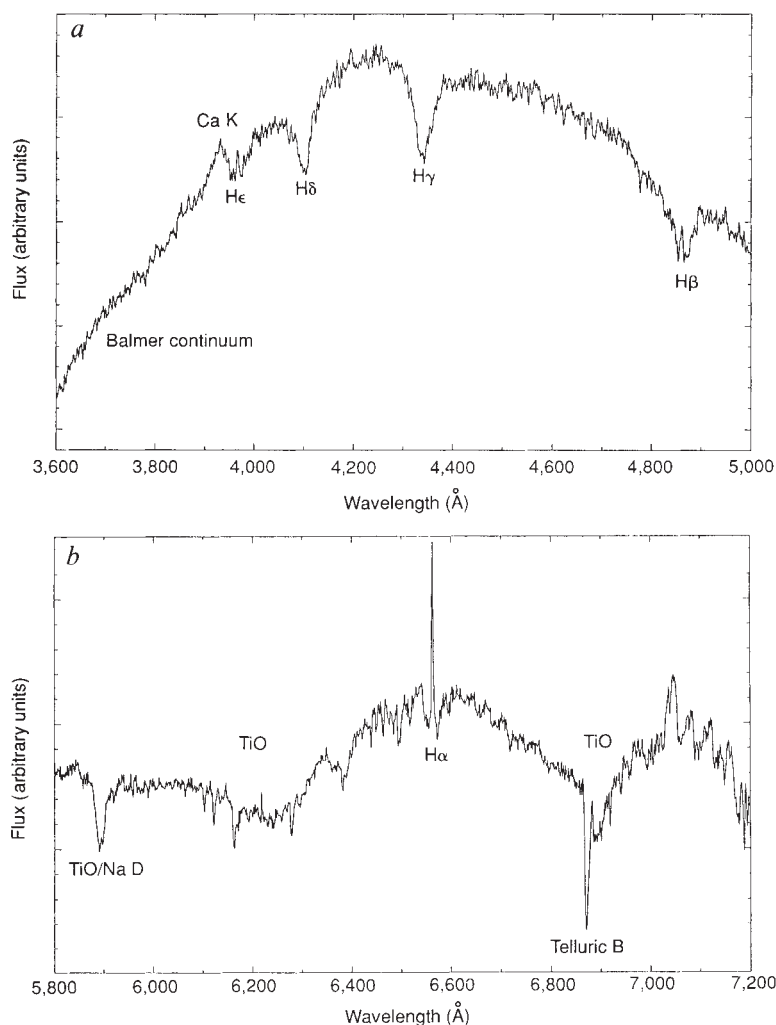


FIG. 2 *a*, Low-resolution (~ 5 Å) blue spectrum (3,600–5,000 Å) of object 1 in the error box of RE1629+781 obtained with the ISIS double spectrograph on the WHT. *b*, Similar-resolution red spectrum (5,800–7,000 Å) obtained with the ISIS double spectrograph on the WHT. The spectra in *a* and *b* are not flux-calibrated.

for the EUV source Feige 24 (ref. 6), which is classified as a DA+dm1–2 binary, leads us to believe that object 1 resembles Feige 24 and is the optical counterpart of RE1629+781. The strong variability of the Balmer emission lines on a timescale of a day reported⁵ for Feige 24 was confirmed by Thorstensen *et al.*² who established the existence of a 4.23-day period. A similar search for an optical period in RE1629+781 is currently in progress.

Comparing the red spectrum with standard spectra (for example, see Jacoby *et al.*⁷) suggests that the cool companion has a spectral type in the range dm2–dm5. To be detected by the WFC, the DA star is expected to have $T_{\text{eff}} > 25,000$ K and, if it is to lie within the normal DA temperature range, $T_{\text{eff}} \leq 60,000$ K. This implies that $9.0 < M_V < 10.0$. We measure $m_V = 13.0$ for RE1629+781 so if the companion is a dm2 star ($M_V = 10.0$)⁸ the distance range is 56–75 pc. For a cool star as late as dm5 ($M_V = 13.5$)⁸ the distance range reduces to 40–63 pc. With these uncertainties, we obtain $13.0 < m_V < 13.75$ for the white dwarf.

To complement the count rates in the WFC S1 and S2 bands (1.1 ± 0.1 s⁻¹ and 0.9 ± 0.3 s⁻¹ respectively) we have obtained a count rate (0.73 ± 0.19 s⁻¹) for the source in the energy range 0.2–3.5 keV from the recently available Einstein Observatory imaging proportional counter (IPC) Slew Survey Catalogue⁹. These three measured count rates can be compared with those obtained by folding a series of model DA spectra (see, for example, ref. 10), normalized to the chosen V magnitude of the white dwarf, through the appropriate instrument response of the three bands in which the measurements were made. We assume that the DA atmospheres are composed of hydrogen

and a trace of helium homogeneously mixed. For a given composition (He/H ratio), we generated curves of constant count rate as a function of stellar temperature and line-of-sight column density, n_H , which corresponded to the observed rates. This is illustrated in Fig. 3 for a pure hydrogen atmosphere where each group of three curves (observed count rate and error limits) corresponds to one filter/instrument combination as labelled. Where all three sets of contours intersect, the model is consistent

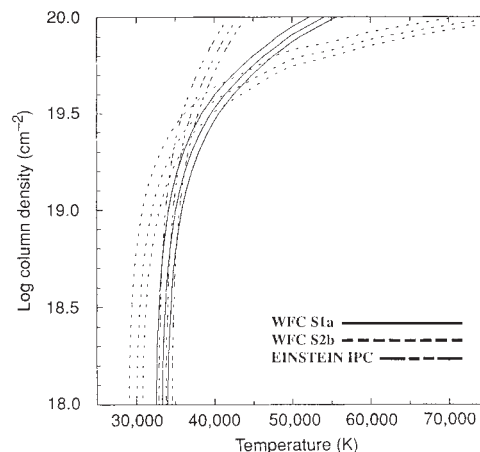


FIG. 3 Curves of constant, measured count rate in particular instrument/filter combinations in the line-of-sight column-density/temperature plane for RE1629+781. This model of a white-dwarf atmosphere is for pure hydrogen. The errors are $\pm 1\sigma$.

TABLE 1 Results of model white-dwarf atmospheres for RE1629 + 781

Composition (He/H)	Temperature range (K)		n_{H} Range (10^{19} cm^{-2})	
	$V=13.0$	$V=13.75$	$V=13.0$	$V=13.75$
0	35,500–37,800	35,900–37,800	1.95–2.80	1.95–2.60
10^{-5}	36,400–37,600	36,400–37,300	1.70–2.30	1.80–2.30
3×10^{-5}	36,400–37,800	40,000	1.00–1.50	1.20
10^{-4}		No fit, therefore excluded.		
10^{-3}		No fit, therefore excluded.		

with the data (as in Fig. 3), and an allowed range of temperature and column density can be determined. Otherwise the model can be excluded. We have followed this procedure for compositions ranging from pure hydrogen to a He/H ratio of 10^{-3} using also the two extremes of V magnitude derived above to normalize the model spectra. The results are summarized in Table 1.

We conclude that the white dwarf has $35,500 \text{ K} < T_{\text{eff}} < 40,000 \text{ K}$, which implies $M_{\text{V}} \approx 9.5$. The abundance of helium (by number) in the star's atmosphere is less than 10^{-4} . Our He/H ratio is similar to values determined by Exosat measure-

ments¹¹ on DA white dwarfs with $T_{\text{eff}} < 45,000 \text{ K}$. We have assumed that the dM star makes negligible contribution to the EUV flux from the system. This is reasonable as the WFC count rate from the very active binary dMe flare star BY Dra (M.A.B., manuscript in preparation) placed at 43 pc, our closest estimate for RE1629 + 781, would be less than 1% of the total observed.

Although they have similar optical spectra, the Feige 24 DA + dM binary system and RE1629 + 781 show different characteristics in the EUV. Observations^{12,13} of Feige 24 have shown a sharp fall-off of flux at the short-wavelength end of the EUV spectrum, but no evidence of the absorption edge at $228 \text{ \AA}$ that would result if helium was a principal source of opacity. The lack of strong absorption features in the EUV spectrum and the presence of narrow photospheric lines in the ultraviolet spectrum prompted Vennes *et al.*¹³ to model the atmosphere of Feige 24 successfully as hydrogen-rich with small traces of 10 elements chosen to have photoionization edges in the soft X-ray and EUV bands. If such a source of opacity exists in RE1629 + 781, then the temperature would have to be significantly higher than we deduce from our models. Alternatively, and interestingly, the atmospheric compositions may be quite different. □

Received 21 June; accepted 8 October 1991.

1. Wells, A. *et al.* *Proc. SPIE* **1344**, 230 (1990).
2. Thorstensen, J. R., Charles, P. A., Margon, B. & Bowyer, S. *Astrophys. J.* **233**, 260–265 (1978).
3. Trümper, J. *Phys. Scripta* **T7**, 209 (1984).
4. Margon, B. *et al.* *Astrophys. J.* **209**, 525–535 (1976).
5. Liebert, J. & Margon, B. *Astrophys. J.* **216**, 18–22 (1977).
6. Margon, B. *et al.* *Astrophys. J.* **210**, L79–82 (1976).
7. Jacoby, G. H., Hunter, D. A. & Christian, C. A. *Astrophys. J. Suppl.* **56**, 257–281 (1984).

8. Straizys, V. & Kuriliene, G. *Astrophys. Space Sci.* **80**, 353–368 (1981).
9. Plummer, D., Schachter, J., Garcia, M. & Elvis, M. *The EINSTEIN Observatory IPC Slew Survey: FITS A3D/CD-ROM Version* (Smithsonian Astrophysical Observatory, 1991).
10. Barstow, M. A. *Mon. Not. R. astr. Soc.* **243**, 182–191 (1990).
11. Paerels, F. B. S. & Heise, J. *Astrophys. J.* **339**, 1000–1012 (1989).
12. Paerels, F. B. S., Bleeker, J. A. M. & Heise, J. *Astrophys. J.* **309**, L33–37 (1986).
13. Vennes, S., Chayer, P., Fontaine, G. & Wesemael, F. *Astrophys. J.* **336**, L25–28 (1989).
14. Bunclark, P. S. & Irwin, M. J. *Proc. Int. Colloq. on Statistical Methods in Astronomy. ESA Spec. Publ. no. 201* (ed. Rolfe, E.) (European Space Agency, 1984).

Surface catalysis studied by *in situ* positron emission

G. Jonkers, K. A. Vonkeman, S. W. A. van der Wal & R. A. van Santen

Koninklijke/Shell-Laboratorium, Amsterdam, PO Box 3003, 1003 AA Amsterdam, The Netherlands

A COMPLETE understanding of heterogeneous catalytic processes requires quantitative *in situ* information on the concentrations of reactants present on the catalyst surface, but few techniques are available to supply this information. Here we show that positron-emitter labelling and scanning, using the isotopes ^{11}C , ^{13}N and ^{15}O , can be used to study the reactions taking place during the catalytic conversion of automotive exhaust. By introducing small pulses of labelled molecules into a reactant stream passing through the catalyst, *in situ* quantitative information on the concentrations and residence times of reactants in the reactor is obtained. The labels are detected using a positron camera, an imaging device adapted from nuclear medicine, and the recorded data are presented as 'reaction images', showing quantitatively the distribution of the label in the catalyst bed as a function of position and time. These data can then be used to quantify reaction kinetics by serving as the input to mathematical simulations based on elementary reaction steps.

Positron-emission computerized tomography (PET) is an imaging technique originating from nuclear medicine^{1,2}, where it is mainly applied as an *in vivo*, noninvasive medical research tool for the investigation of brain and heart diseases. PET has been applied industrially^{3,4} by labelling the phase or particle of interest with a positron-emitting nuclide; examples are monitoring (nonreactive) flow of water in oil reservoir rock³ and establishing flow profiles in extruders⁴. The availability of positron-emitting nuclides of carbon (^{11}C , half-life $t_{1/2} = 20.4 \text{ min}$), nitrogen (^{13}N , 9.96 min) and oxygen (^{15}O , 2.07 min) allows labelling of many reactant molecules. Because of these

short half-lives, PET cameras are located near a radionuclide production facility, usually a cyclotron. Because of the high specific activity and the fact that individual molecules can be detected, experiments can be carried out using only minute amounts of labelled molecules (of the order of 0.1 pmol).

Because ^{11}C -, ^{13}N - or ^{15}O -labelled molecules are chemically indistinguishable from unlabelled ones, they offer the possibility of quantitatively measuring the presence and residence times of actual reactants on a catalyst surface under operating conditions (*in situ*). The ^{11}C -, ^{13}N - or ^{15}O -nucleus decays by emitting a positron, the antimatter counterpart of an electron. On an encounter between a positron and an electron, these two particles are instantaneously converted (annihilated) into two 511-keV γ -photons with high penetrating power, allowing noninvasive measurements. Simultaneous detection of both γ -photons by two individual detectors circuited in coincidence^{2,5} defines the line on which the annihilation occurred (Fig. 1). The use of a PET camera as detection system should allow absolute positron-emitter concentrations to be monitored at a well-defined time and location, and thus allow 'imaging' of catalytic interactions. By using labelled molecules such as ^{11}CO , C^{15}O , ^{15}OO , $^{11}\text{CO}_2$, C^{15}OO or ^{13}NN and a PET camera, we have demonstrated this technique in the investigation of automotive exhaust catalysis reactions.

A Pt-CeO₂/γ-Al₂O₃ exhaust catalyst was placed in a tubular plug-flow reactor ($d = 7 \text{ mm}$) and continuously exposed to synthetic exhaust gas until steady-state conditions were obtained (Table 1). The conditions were chosen in such a way that the conversion of CO with O₂ into CO₂ was exclusively limited by intrinsic reaction kinetics and not by transport phenomena. The reactor was positioned horizontally in the centre of a NeuroECAT[®] PET instrument^{6,7} (Fig. 1). A minute amount of ^{11}CO was added as a pulse to the synthetic exhaust gas passing through the reactor without disturbing the steady state of the catalyst (Table 1). The tomograph was operated in such a way that only events that were coincident at detectors in both opposing banks parallel to the flow tube were registered (Fig. 1). To obtain the optimal spatial resolution, we analysed the recorded

data by selecting either coincident events between directly opposite detectors (11 positions in total) or (average) coincidences between two opposite pairs of adjacent detectors viewing a single point on the reaction tube axis (10 positions). In this way the positron-emitter concentration at 21 equidistant ($\Delta d = 1.1$ cm)^{6,7} positions over the reactor was measured simultaneously, and this measurement could be repeated every 1.2 s.

The recorded data, which have been corrected for radioactive decay, attenuation effects and detector efficiencies, are displayed in a 'reaction image' (Fig. 2) showing horizontally the axial concentration profile of the ^{11}C -label over the reactor. The vertical axis represents the time coordinate of the reaction image. The magnitude of the signal is represented in colours derived from a linearly scaled colour bar. As soon as the ^{11}CO pulse reaches the catalyst bed, it is adsorbed and can thereafter be converted into $^{11}\text{CO}_2$. The $^{11}\text{CO}_2$ is desorbed and carried away with the gas phase towards the outlet of the reactor. The $^{11}\text{CO}_2$ interacts with the ceria (CeO_2) on the catalyst surface through an adsorption/desorption process and is thereby delayed (see below).

The analysis of the recorded data is facilitated by drawing cross-sections through the images. This may be done either at a defined time, resulting in the ^{11}C -label distribution over the length of the reactor (at $t = 79$ s; PRF in Fig. 2), or at a defined position in the reactor, leading to the residence time distribution of the ^{11}C -label (at $x = -5.5$ cm; RTD in Fig. 2).

Under identical steady-state conditions (see Table 1) in separate experiments, six different labelled compounds were pulsed into the synthetic exhaust gas stream (Fig. 3a-f). Figure 3a demonstrates the plug-flow operation of the reactor when an inert gas $^{13}\text{N}_2$ is led over an inert carrier (SiC). There is a considerable difference in residence time between the reaction images recorded for the $^{11}\text{CO}_2$ and C^{15}OO labelling experiments (Fig. 3b and e respectively) as well as between those for ^{11}CO and C^{15}O (Fig. 3c and f).

Comparison of reaction images 3b and e indicates that the C atoms of the CO_2 molecules entering the reactor have all left the catalyst bed before their original O atom partners arrive at the reactor outlet, demonstrating their separation during the reaction process. From these and other experiments⁸, it was concluded that CO_2 molecules adsorb and decompose on the ceria surface. The ceria surface easily forms carbonate groups⁹⁻¹¹, and we therefore propose the formation of these species as an explanation. Under the reaction conditions, the cerium carbonate groups can decompose again to form CO_2 and ceria, where the O atoms of the desorbing CO_2 are not necessary identical to those of the originally adsorbed CO_2 .

TABLE 1 Typical reaction and measuring conditions

Composition of synthetic exhaust gas	88.5 vol % Ar 10.0 vol % CO_2 1.0 vol % CO 0.5 vol % O_2
Flow rate	$30 \mu\text{mol s}^{-1}$
Catalyst system	3.9 g Pt-CeO ₂ /γ-Al ₂ O ₃ (30–80 mesh)
Platinum loading	0.12 wt % Pt
Temperature	140 °C
Pressure	1 bar
Conversion	76 %
Injected pulse	
Total volume	200 μl (carrier added)
Cylindrical shape length	2 cm
Duration	0.4 s
Activity (^{11}C)	50 MBq (2 μCi)
Number of molecules (^{11}C)	0.15 pmol (carrier free)

The difference between the ^{11}CO and the C^{15}OO labelling experiments (Fig. 3c and f) is due to the same effect. CO is oxidized on the platinum surface, and the CO_2 that is formed interacts with ceria. The bright narrow band at the bottom of the reaction image for ^{15}OO labelling experiments (Fig. 3d) is comparable to that of Fig. 3a and represents ^{15}OO molecules that do not interact with the catalyst bed. The remaining ^{15}OO molecules adsorb and dissociate on the Pt surface, after which they react with adsorbed CO forming C^{15}OO . The C^{15}OO desorbs from the Pt surface and becomes adsorbed on the ceria surface as described above. This causes the smeared activity distribution in the reaction image. The irreversible dissociative nature of the O_2 adsorption is demonstrated by product identification of the ^{15}O -labelled species leaving the catalyst bed. In the initial peak, consisting of ^{15}O -labelled compounds that passed through the catalyst bed without interaction, only ^{15}OO molecules were present, whereas the remaining ^{15}O -atoms all left the catalyst bed in C^{15}OO molecules.

This positron-emitter labelling and scanning technique has been applied to an effectively one-dimensional (spatial) reactor. Here both the spatial (9 mm) and temporal (1.2 s) resolution achieved are dictated by the technical design of the NeuroECAT[®], which was developed around 1980. Specially designed detector arrays may achieve a temporal resolution in the millisecond region. Future experiments may benefit from the progress made in the design of PET cameras, allowing monitoring of the three-dimensional distribution of ^{11}C -, ^{13}N - or ^{15}O -labelled molecules with a resolution close to its physical

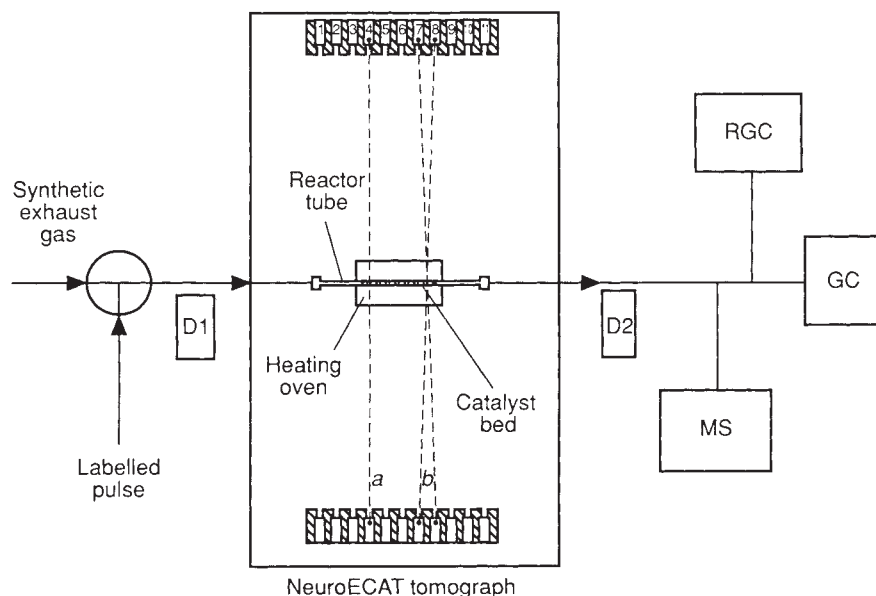


FIG. 1 Schematic drawing of the experimental setup. D1 and D2 are separate radiation detectors for accurate monitoring of the inlet and outlet of the reactor system. The NeuroECAT[®] is a PET tomograph (technology from about 1980) with eight detector banks, each containing 11 individual detectors, arranged in an octagon. The detectors of the two banks parallel to the flow tube are drawn to scale. Dashed lines interconnect directly opposing (a) or opposing adjacent (b) detectors. Outlet gases from the reactor are analysed by a gas chromatograph (GC), a radio gas chromatograph (RGC) and a mass spectrometer (MS).

FIG. 2 *a*, Reaction image of a ^{11}CO experiment (see Table 1). The horizontal scale gives the position along the catalyst bed. The yellow dotted lines indicate the position of the catalyst bed. The vertical scale indicates the residence time. The colours indicate the ^{11}C -label concentration according to the linearly scaled colour bar. Also indicated are a profile (PRF) of the distribution of the label along the reactor axis (at time $t = 79$ s), and a residence time distribution (RTD) of the label (at position $x = -5.5$ cm). *b*, Computer simulation of the reaction image recorded for the ^{11}CO experiment.

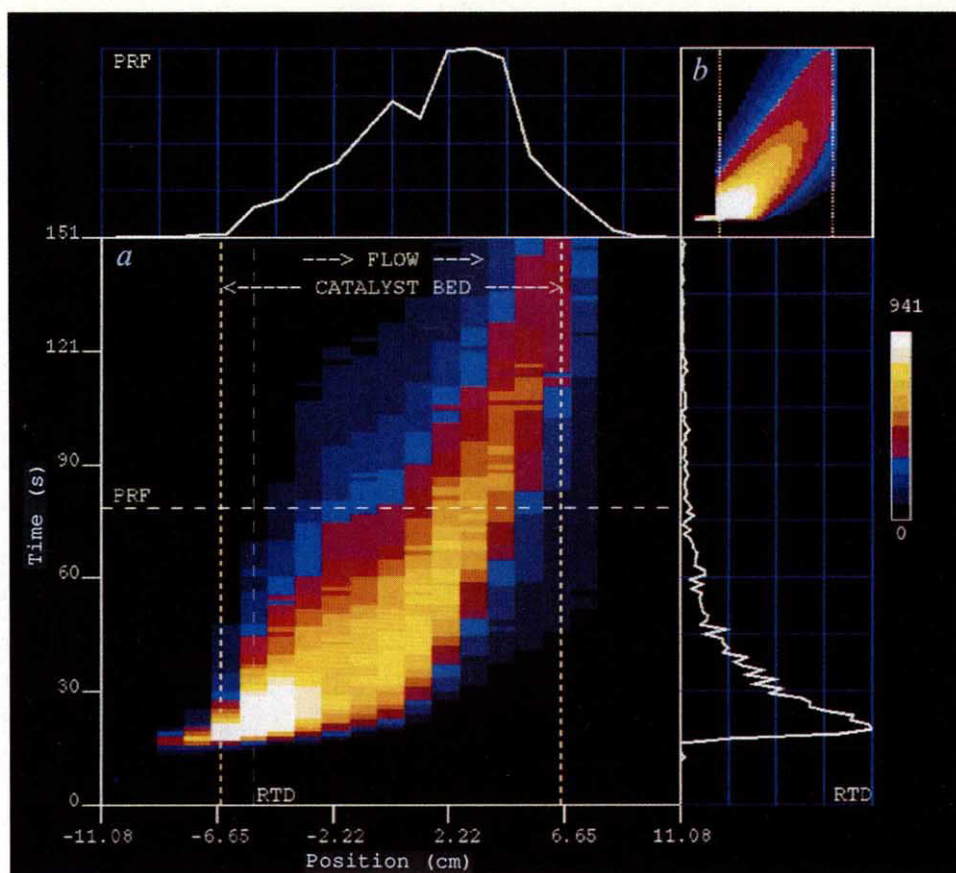
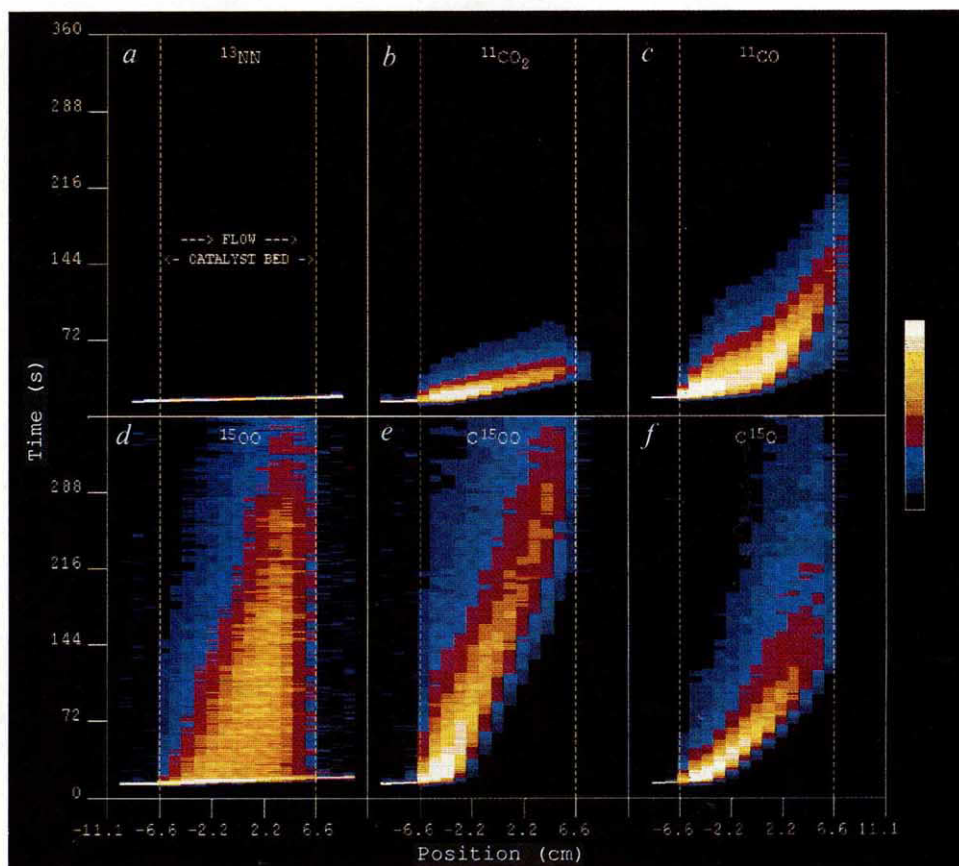


FIG. 3 Six reaction images of separate pulse experiments with different labels under identical conditions (see Table 1). *a*, ^{13}NN . The labelled molecules pass unhindered through the catalyst bed, showing their inertness. *b*, $^{14}\text{CO}_2$. The CO_2 molecules are delayed by the sorptive interaction with ceria on the catalyst surface. *c*, ^{11}CO . The delay of the CO molecules caused by adsorption on the Pt surface is followed by reaction to form CO_2 , CO_2 desorption and the sorptive interaction between CO_2 and the ceria. *d*, ^{15}OO . Some of the labelled molecules pass unhindered through the catalyst bed (bright 'line' at bottom); the others are adsorbed on the platinum and react to form CO_2 , which desorbs and becomes adsorbed on ceria with exchange of O atoms. This clearly shows the irreversible and dissociative nature of the oxygen adsorption on platinum. *e*, C^{15}OO . Interaction and O exchange between CO_2 and ceria. *f*, C^{15}O . CO oxidation on Pt, followed by CO_2 desorption and sorptive CO_2 /ceria interaction with the O-atom exchange.



limit (2 mm)^{1,2}, ultimately dictated by the mean free path of the positron before annihilation.

The technique is not limited to automotive exhaust catalysis research. In homogeneous as well as in heterogeneous gas- or liquid-phase catalysis, both intrinsic kinetics and transport phenomena can be studied. The amount of data gathered in each experiment is very large and can therefore allow accurate quantification of reaction kinetics if used as the input to mathematical simulations rigorously based on elementary reaction steps. The simulation of the ¹¹CO pulse experiment (Fig. 2b) is the result of such a modelling study, which was based on kinetic constants for oxidation of CO obtained in surface science studies^{8,12,13}. The labelled molecules used in the experiments are more or less directly available from the irradiated target. For this technique to be more widely applicable, labelling routes to basic reactants (for example, alkanes and aromatics for the petrochemical industry) have to be developed. Where *in situ* information from actual operating conditions is required, it should become a very useful technique. □

Received 15 July; accepted 14 October 1991.

1. *Positron Emission Tomography and Autoradiography—Principles and Applications for the Brain and Heart* (eds Phelps, M. E., Maziotto, J. C. & Schelbert, H. R.) (Raven, New York, 1986).
2. *The Physics of Medical Imaging* (ed. Webb, S.) (Adam Hilger, Bristol, 1988).
3. Van den Bergen, E. A., Jonkers, G., Strijkmans, K. & Goethals, P. *Int. J. Radiat. appl. Instrum. E: Nucl. Geophys.* **3**, 407–418 (1989).
4. Hawkesworth, M. R., Bemrose, C. R., Fowles, P. & O'Dwyer, M. A. in *Tomography and Scatter Imaging IOP Short Meetings Series 19* (eds MacCuaig, N. & Holt, R.) 67–79 (Institute of Physics, Bristol, 1989).
5. Knoll, G. F. *Radiation Detection and Measurement* (Wiley, New York, 1989).
6. Williams, C. W. *et al.* *IEEE Trans. nucl. Sci.* **NS-28**, 1736–1740 (1981).
7. Hoffman, E. J., Phelps, M. E. & Huang, S. C. *J. nucl. Med.* **24**, 245–257 (1983).
8. Vonkeman, K. A. thesis, Technical Univ. Eindhoven (1990).
9. Jin, T., Okuhara, T., Mains, G. J. & White, J. M. *J. phys. Chem.* **91**, 3310–3315 (1987).
10. Daniel, D. W. *J. phys. Chem.* **92**, 3891–3899 (1988).
11. Li, C. *et al.* *J. chem. Soc. Faraday Trans. 1* **85**, 929–943 (1989).
12. Engel, T. & Ertl, G. in *Adv. Catal.* **28** (eds Eley, D. D., Pines, H. & Weisz, P. B.) 2–78 (1979).
13. Lynch, D. T., Emig, G. & Wanke, S. E. *J. Catal.* **97**, 456–468 (1986).

ACKNOWLEDGEMENTS. We thank J. de Jong and H. Oosterbeek for their experimental assistance and K. Strijkmans, P. Goethals, I. Lemahieu and K. de Kesel, Institute for Nuclear Sciences, Rijksuniversiteit Ghent, Belgium for production of the radioisotopes, synthesis of the labelled reactants and help in experimenting with the NeuroECAT[®].

Femtosecond laser control of a chemical reaction

E. D. Potter, J. L. Herek, S. Pedersen, Q. Liu & A. H. Zewail

Arthur Amos Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, California 91125, USA

THE critical stage in a chemical reaction—the progression through the transition state from reagents to products—occurs in less than a picosecond (10^{−12} s). Using laser pulses of femtosecond (10^{−15} s) duration it is possible to probe the nuclear motions throughout formation and break-up of the transition state^{1,2}. The coherence and very short duration of these femtosecond pulses provides a means to influence the course of the reaction during this stage if the time resolution is made sufficiently short. Here we describe a demonstration of such control of a chemical reaction on the femtosecond timescale. Using two sequential coherent laser pulses, we can control the reaction of iodine molecules with xenon atoms to form the product XeI by exciting the reactants through the transition state, in a two-step process. The yield of product XeI is modulated as the delay between the pulses is varied, reflecting its dependence on the nuclear motions of the reactants.

For the most general elementary chemical reaction,



the products (AB + C) are formed from reagents (A + BC) after passing through the transition state [A⋯B⋯C][‡]. This state, a

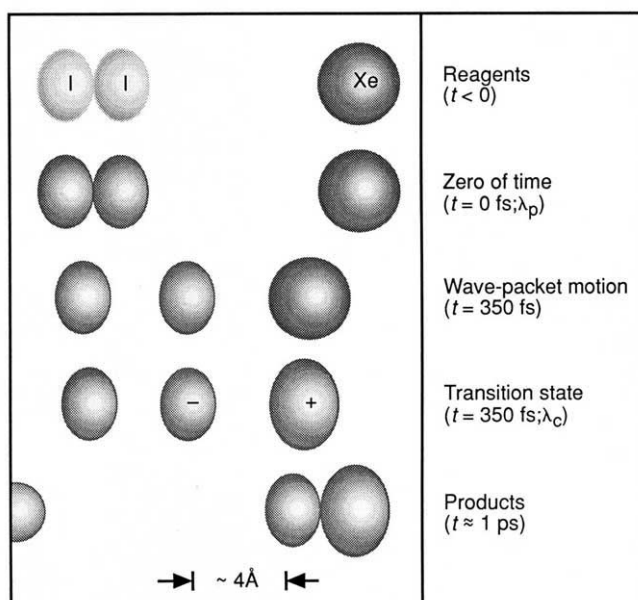
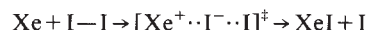


FIG. 1 Snapshots of the reaction $A + BC \rightarrow AB + C$ at different delay times. λ_p and λ_c refer to the pump and control wavelengths of the femtosecond pulses. The transition state for this case of $Xe + I_2$ is formed by electron harpooning from Xe to I_2 . The dynamics of the reaction is complete in ~ 1 ps. The reaction is illustrated here for a particular internuclear configuration, without details of the pathway (see text), for simplicity.

'collision complex', has six internal degrees of freedom, and energy can flow among them in a short time, typically picoseconds or less. At longer times, the distribution of energy becomes randomized and the system may reach a statistical limit in which the energy is shared among all degrees of freedom. If the energy is supplied to the reactants over a short timescale and further chemical change is induced after a prescribed time delay, it may be possible to control the trajectory of the nuclear motions towards product formation, either by the direct or the indirect formation of the transition state.

Here we consider the bimolecular (atom-molecule) reaction of rare gas atoms (Xe) and halogen molecules (I_2). The reaction involves the translational motion of Xe towards I_2 , the nuclear vibrational motion of iodine, and the formation of the transition state by Xe harpooning the I_2 through electron transfer:



The cross-section for this type of reaction is generally very large (~ 100 – 200 Å²), and is essentially determined by the critical distance at which the harpooning occurs, which is typically larger than the van der Waals radius. Following the harpooning, the Coulomb interaction leads to the formation of XeI and the separation of product XeI from the iodine atom. This harpooning mechanism, introduced by Polanyi³, is important for a variety of reactions in the gas phase, on surfaces and in the condensed phase^{4–6}. For the rare gas/halogen systems, new studies have shown the importance of the harpooning mechanism in laser-assisted reactions (for reviews, see refs 7, 8).

The idea here is to exploit the nuclear motions (Fig. 1) on a femtosecond timescale. We use two femtosecond pulses separated by a time delay to pump and control the $Xe + I_2$ reaction. The first pulse, λ_p , creates a well defined, coherent wavefunction⁹ (or more accurately, a wave packet) in an intermediate state, the B state of iodine. The wave packet moves⁹ back and forth with a well defined period as the internuclear separation in the iodine molecule varies between 2.5 and 5 Å (Fig. 2). Some femtoseconds after the pump pulse, a second pulse λ_c is sent to lift this wave packet above the reaction threshold for reaction with Xe. Previous studies¹⁰ have shown that this reaction involves the collision pair $I_2 \cdots Xe$ (see below). If, for example, λ_c is

applied when the packet is at the point corresponding to the appropriate internuclear separation, this pulse can take the packet into the region where the harpoon mechanism takes place. The energy of the two pulses exceeds the reaction threshold^{11,12} ($52,460 \pm 400 \text{ cm}^{-1}$; $191 \pm 2 \text{ nm}$) by $\sim 5,000 \text{ cm}^{-1}$. Formation of the XeI product can therefore be switched on or off depending on the position of the wave packet when λ_c is applied; that is, on the time delay between λ_p and λ_c . One therefore expects to see a modulation of the yield of XeI, as the delay is increased, which is in phase with the signal from the B state of I_2 . This is indeed what is observed experimentally (Fig. 2).

The experiments were done by irradiating a mixture of Xe and I_2 (10–300 torr and 1 torr respectively) by the two femtosecond pulses. The product yield was identified by the chemiluminescence of the XeI ($\text{B} \rightarrow \text{X}$) transition near 253 nm. The laser apparatus used in these experiments has been described in detail elsewhere^{1,2}. The control pulse was delayed relative to the pump using a Michelson interferometer before entering the reaction chamber.

Figure 2 shows the yield of XeI as a function of time, and Fig. 3 gives the chemiluminescence spectra of XeI at four time delays between the two pulses. In addition to product emission near 253 nm, we observe a signal characteristic of bare iodine (340 nm), which is used here for internal calibration and has been assigned previously to the $\text{D}' \rightarrow \text{A}'$ transition. To confirm our assignment, we compared the results of experiments using bare iodine, iodine with 50 torr Ar, and iodine with 50 torr Xe. The bare iodine experiment showed no fluorescence from XeI in this region, and the bare iodine signal at 340 nm was weak. The argon/ I_2 experiment showed a strong signal at 340 nm (stronger than the 253 nm signal in the Xe cell), but gave virtually no signal at 253 nm (for a scale expanded by a factor of 20, a small signal at this wavelength was observed, which can be attributed to the expected I_2^{**} fluorescence at $\sim 280 \text{ nm}$). Argon is known to quench the excited-state iodine into the D' state, which explains the strong 340-nm $\text{D}' \rightarrow \text{A}'$ fluorescence in this experiment. The Xe/ I_2 experiment, on the other hand, showed

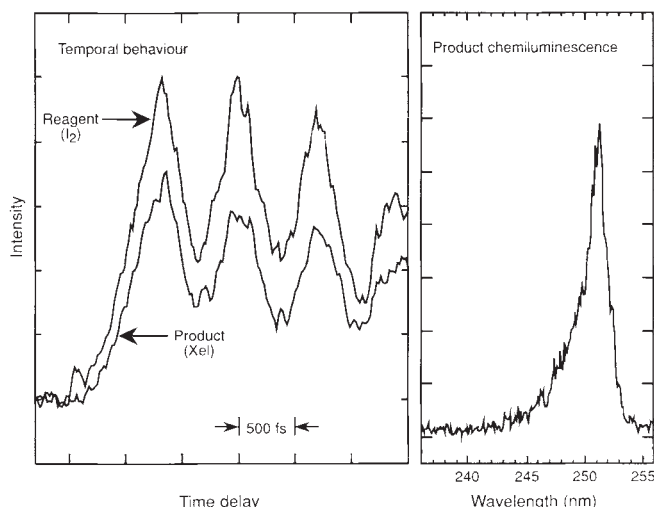


FIG. 2 Left: Femtosecond temporal behaviour of product XeI chemiluminescence (at 253 nm) as a function of the delay time between pump and control pulses (solid line). For reference, the change with time of reagent I_2 (at 340 nm) is also shown (dotted line); $\lambda_p \sim 515 \text{ nm}$ and $\lambda_c \sim 260 \text{ nm}$. Right: Product XeI chemiluminescence spectrum at $\sim 253 \text{ nm} \pm 2 \text{ nm}$. The two femtosecond pulses were selected^{1,2} from two white-light continua, and one (the control pulse) was reamplified using the third harmonic of a Nd/YAG laser and then doubled in a KDP crystal^{1,2}. In this case, the pulse duration was typically $\sim 150 \text{ fs}$, larger than the characteristic 50–70-fs pulses of the apparatus^{1,2}.

a very strong signal at 253 nm (at 50 torr pressure), as well as the signal at 340 nm.

To demonstrate the effect more clearly, we display in Fig. 3 the XeI chemiluminescence yield at four delay times between the pump and control pulses. At negative time (A), the control pulse arrives before the pump pulse and we observe no XeI emission. For simultaneous pump and control pulses (B), there

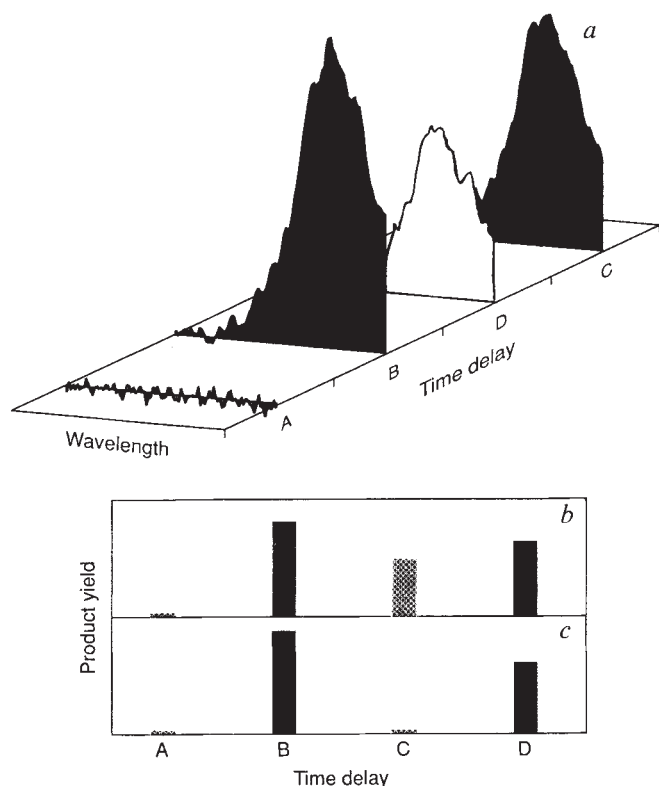
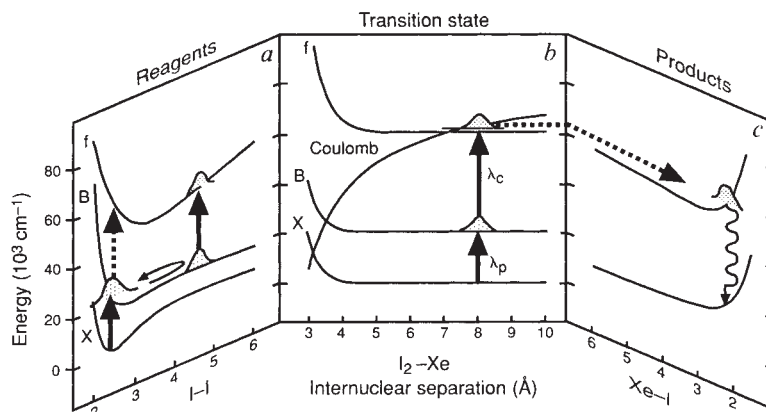


FIG. 3 *a*, Experimental chemiluminescence spectra of product XeI (at 253 nm; low resolution) as a function of delay time between pump and control femtosecond pulses. A (negative time); B (zero of time); C (350 fs); D (700 fs). Unlike the spectrum in Fig. 2, the resolution here is decreased (11 nm) to enhance the signal-to-noise ratio. *b*, Experimental bar graph representing the 253-nm signal intensity at the time delays A, B, C and D. *c*, The expected bar graph of the 253-nm signal intensity for a pulse width much smaller than the vibrational period at this total energy.

FIG. 4 Three sections through the potential energy surface for the reaction $\text{Xe} + \text{I}_2 \rightarrow \text{XeI} + \text{I}$. *a*, The potential along the I—I vibrational coordinate (large separation from Xe). Three potential-energy curves are shown, for the I_2 ground state (X), excited state (B) and ion-pair state (f). *b*, The potential, along the Xe— I_2 coordinate. *c*, The potential along the Xe—I vibrational coordinate (large separation from other I). The wave packets are shown to illustrate the pathway for the nuclear motion towards product formation. Panel *a* shows pumping and control via the B state, while the Xe atom is at a large distance (harpooning occurs when Xe collides with I_2 in the ion-pair state). The wave packet can be lifted up to the ion-pair state at either of the turning points, depending on the choice of λ_c . Panel *b* shows the pump and control of the 'collision pair' of $\text{I}_2 \cdots \text{Xe}$ directly to the harpooning region. The pump wavelength matches a span of $\text{I}_2 \cdots \text{Xe}$ separations and is shown here for only one particular case. Panel *c* shows the formation of product XeI from the harpoon region.



is a maximum in the product yield, whereas for a time delay of 350 fs (C) the yield decreases again. The pulse width used here does not allow us to observe a complete modulation of the yield, but this should be easily achieved with shorter pulses (Fig. 3).

The use of two femtosecond pulses for pump and control defines the zero of time, and the 'switch' control effect is then the result of the coherent motion of the iodine wave packet along the I—I coordinate. Control of this coherence is reminiscent of the problem of mode-selective chemistry. Following experimental observations¹³, it was proposed some time ago that it may be possible to localize excitation in a given bond or mode by using ultrashort pulses that are capable of forming coherent superposition of states¹⁴. Tannor and Rice¹⁵ have proposed a scheme for selective photochemistry with ultrashort pulses using an illuminating wave packet description. The experiments reported here should have extension to studies of mode-selective chemistry¹⁶ and to systems of the type proposed by Tannor and Rice.

Several other schemes have been proposed. Mukamel¹⁷ has considered the effect of intense pulses as a way of controlling relaxation. Slewa *et al.*¹⁸ have demonstrated locking of relaxation, using multiple-pulse excitation, and proposed extending the method to large systems. Brumer and Shapiro¹⁹ have considered theoretically the importance of the phase difference between two continuous-wave laser sources in driving different product channels (the eigenstate picture). Rabitz and co-workers²⁰ have proposed the use of optimally shaped pulses, and Manz and colleagues²¹ have suggested other new control systems for femtosecond chemistry. There have been recent successful studies^{22–25} of selectivity (and enhancement) which use selective laser excitation without temporal resolution. The temporal resolution of the nuclear motion^{1,2,9}, the population²⁶ and phase²⁷ selectivity, and the chemical control reported here suggest a number of new experiments. For example, very recently, the Rice group²⁸ has considered theoretically the problem of three laser fields in diatomics, and we should be able to test experimentally the degree of control (contrast ratio for product) in these interesting diatomic systems and also in reactions on ground-state potential surfaces.

The time resolution should also be helpful in studying the mechanism, as we hope to explore in future experiments. Setser's finding^{8,10} indicates that the $\text{Xe} + \text{I}_2$ reaction takes place via collision pairs, and thus our control is exercised 'during' the collision. Another reaction channel, however, is likely to be involved also, depending on the λ_p , λ_c combinations. Following the creation of the packet in the B state, the control pulse can promote the wave packet to the ion-pair state of I_2 which then collides with Xe to form XeI . Donovan *et al.*^{11,12}, in their work on I_2 , have demonstrated this reactive quenching by Xe. In this case, the timing of the B-state wave packet motion controls the formation of the transition state and the yield of product, even

though the collisional time with Xe is nanoseconds. Figure 4 shows different cuts along three coordinates of the potential for illustration of the two cases. The scheme in the middle panel is supported by the observation of two-colour two-photon assistance of the reaction reported by Setser's group^{8,10} and by the work of Soep and Juvet²⁹ on van der Waals complexes. Determination of the phase of the wave packet motion (relative to yield modulation), tuning of λ_c (to reach different points on the potential), and temporal resolution of the chemiluminescence will establish the nature of the entrance channel at different combinations of λ_p and λ_c .

The experimental results reported here indicate that the product yield of a chemical reaction ($A + BC$) can be controllable temporally if the motion of the wave packet towards (direct or indirect) formation of the transition state is timed using pump-and-control coherent femtosecond pulses. The methodology should have applications to other bimolecular (and unimolecular) reactions and to other systems in reactive collisions (see, for example, refs 30–32). We can also hope to extend these ideas to larger molecular systems. □

Received 13 November; accepted 3 December 1991.

- Zewail, A. H. *Science* **242**, 1645–1653 (1988).
- Khundkar, L. R. & Zewail, A. H. *Ann. Rev. Phys. Chem.* **41**, 15–40 (1990).
- Polanyi, M. *Atomic Reactions* (Williams and Norgate, London, 1932).
- Polanyi, J. C., Xu, G.-Q. & Zeng, H.-C. *Faraday Discuss. Chem. Soc.* **91**, (1991).
- Okada, F., Wiedeman, L. & Apkarian, V. A. *J. Phys. Chem.* **93**, 1267–1272 (1989).
- Levine, R. D. & Bernstein, R. B. *Molecular Reaction Dynamics and Chemical Reactivity* (Oxford University Press, 1987).
- Donovan, R. J. *et al. Faraday Discuss. Chem. Soc.* **84**, 221–238 (1987).
- Qin, J., Nelson, T. O. & Setser, D. W. *J. Phys. Chem.* **95**, 5374–5378 (1991).
- Bowman, R. M., Dantus, M. & Zewail, A. H. *Chem. Phys. Lett.* **161**, 297–302 (1989).
- Qin, J. & Setser, D. W. *Chem. Phys. Lett.* **184**, 121–127 (1991).
- O'Grady, B. V. & Donovan, R. J. *Chem. Phys. Lett.* **122**, 503–506 (1985).
- Donovan, R. J., Holmes, A. J., Langridge-Smith, P. R. & Ridley, T. J. *Chem. Soc. Faraday Trans.* **85**(2), 541–548 (1988).
- Felker, P. M. & Zewail, A. H. *Phys. Rev. Lett.* **53**, 501–505 (1984).
- Bloembergen, N. & Zewail, A. H. *J. Phys. Chem.* **88**, 5459–5465 (1984).
- Tannor, D. J. & Rice, S. A. *Adv. Chem. Phys.* **LXX** (Part I), 441–523 (1988).
- Manz, J. & Parmenter, C. S. *Chem. Phys.* **139**, 1–4 (1989).
- Mukamel, S. & Shan, K. *Chem. Phys. Lett.* **117**, 489–494 (1985).
- Slewa, E. T., Glasbeek, M. & Zewail, A. H. *J. Phys. Chem.* **90**, 1232 (1986).
- Brumer, P. & Shapiro, M. *Chem. Phys. Lett.* **126**, 541–546 (1986).
- Park, S. M., Lu, S.-P. & Gordan, R. J. *J. Chem. Phys.* **94**, 8622–8624 (1991).
- Shi, S., Woody, A. & Rabitz, H. *J. Chem. Phys.* **88**, 6870–6883 (1988).
- Combariza, J. E., Manz, J. & Paramonov, G. K. *Faraday Discuss. Chem. Soc.* **91**, (1991).
- Butler, L., Hints, E. J., Shane, S. F. & Lee, Y. T. *J. Chem. Phys.* **86**, 2051 (1987).
- Vander Wal, R. L., Scott, J. L. & Crim, F. F. *J. Chem. Phys.* **92**, 803 (1990).
- Zhang, L., Fuss, W. & Kompa, K. L. *Chem. Phys.* **144**, 289–297 (1990).
- Park, S. M., Lu, S.-P. & Gordan, R. J. *J. Chem. Phys.* **94**, 8622–8624 (1991).
- Gerdy, J., Dantus, M., Bowman, R. M. & Zewail, A. H. *Chem. Phys. Lett.* **171**, 1–4 (1990).
- Scherer, N. F., Ruggiero, A. J., Du, M. & Fleming, G. R. *J. Chem. Phys.* **93**, 856–857 (1990).
- Armstrup, B., Carlson, R. J., Matro, A. & Rice, S. A. *J. Phys. Chem.* **95**, 8019–8027 (1991).
- Juvet, C., Boivineau, M., Duval, M. C. & Soep, B. *J. Phys. Chem.* **91**, 5416–5422 (1987).
- Barnes, M. D., Brooks, P. R., Curl, R. F. Jr & Harland, P. W. *J. Chem. Phys.* **94**, 5245 (1991).
- Kleiber, P. D., Lyrra, A. M., Sando, K. M., Heneghan, S. P. & Stwalley, W. C. *Phys. Rev. Lett.* **54**, 2003 (1985).
- Breckenridge, W. H., Juvet, C. & Soep, B. *J. Chem. Phys.* **84**, 1443 (1986).

ACKNOWLEDGEMENTS. We thank R. W. Quan for help in preparing samples, and R. Donovan, D. Setser, B. Soep and I. Smith for discussions. This work was supported by the US Air Force Office of Scientific Research and by the NSF.

Impact of aircraft and surface emissions of nitrogen oxides on tropospheric ozone and global warming

Colin Johnson*, Jim Henshaw* & Gordon McInnes†

* Modelling and Assessments Group AEA Environment and Energy, Harwell Laboratory, Didcot, Oxfordshire, OX11 0RA, UK

† Air Pollution Division, Warren Spring Laboratory, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2BX, UK

ACTUAL and potential increases in aircraft traffic are causing concern about the effects of aircraft exhaust emission on atmospheric chemistry. Model results¹⁻³ and measurements⁴⁻⁶ in the Northern Hemisphere have shown that growth in surface emissions of nitrogen oxides and hydrocarbons leads to increases in concentration of tropospheric ozone. Tropospheric ozone is toxic to plants, humans and other organisms, and it is a greenhouse gas⁷⁻⁹. The radiative forcing of surface temperature is most sensitive to changes in tropospheric ozone at a height of ~12 km (ref. 8), where aircraft emissions of nitrogen oxides are at a maximum and where the model sensitivity of ozone to nitrogen oxide emissions is enhanced. Our model results show that the radiative forcing of surface temperature is about thirty times more sensitive to aircraft emissions of nitrogen oxides than to surface emissions. We also find that the impact on global warming of increases in tropospheric ozone caused by increases in surface emissions of nitrogen oxides has previously been overestimated by a factor of five^{1,10}, owing to an error in the calculation of the ozone budget.

Previous two-dimensional global modelling¹¹⁻¹⁶ of the effects of nitrogen oxide NO_x emissions from tropospheric aircraft has led to predictions of increases of 1–15% in ozone in a latitudinal band centred at 45°N when aircraft emissions are included. Much of this variation in ozone increase is explained by differences in emission estimates, and these and other modelling studies are compared in refs 14 and 17. In the stratosphere, models

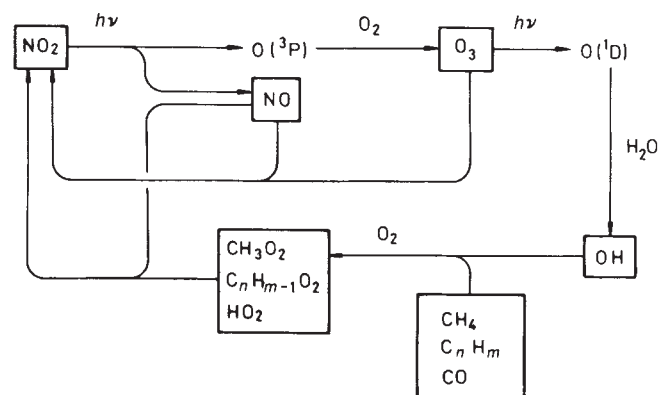


FIG. 1 Summary of the principal chemistry of ozone formation in the troposphere. The inner cycle involves the photodissociation of NO_2 leading to ozone and NO formation, and the subsequent back reaction of these species to regenerate NO_2 . Competing with this back reaction is the outer cycle involving the formation of peroxy radicals from methane, nonmethane hydrocarbons (represented by C_nH_m) and carbon monoxide. The reaction of peroxy radicals with NO forms NO_2 and hence regenerates O_3 . The outer cycle dominates when the NO concentration is $\sim 10^3$ times greater than the ozone concentration as a consequence of the chemical rate coefficients involved. The outer cycle is a net producer of ozone because the reaction of $\text{O}(^1\text{D})$ with H_2O produces two OH radicals. Other chemical losses of ozone also occur¹⁸, and these include reactions with OH and HO_2 radicals as well as oxidation of NO to form NO_2 where the NO_2 is not subsequently photolysed to form $\text{O}(^3\text{P}) + \text{NO}$.

predict decreases in ozone when projected aircraft NO_x emissions are included¹⁸.

Here we use a two-dimensional (latitude–altitude) model of the global troposphere to 24 km to predict changes in ozone concentrations, and those of other gases, in response to changes in NO_x emissions. The model takes into account the emissions, chemistry, transport and deposition of tropospheric trace gases. A complete description of the model, including comparisons with observations, is given elsewhere^{19,20}. The chemical mechanisms of ozone production and destruction are summarized in Fig. 1.

A baseline case in which there were no aircraft emissions was compared with two situations in which we estimated the aircraft emissions for the years 1990 and 2000, and with a sensitivity study in which surface NO_x emissions from man-made sources were increased by 10%. Details are given in Fig. 2a and b, which shows the latitudinally averaged changes to the ozone concentrations in the Northern and Southern Hemispheres when these

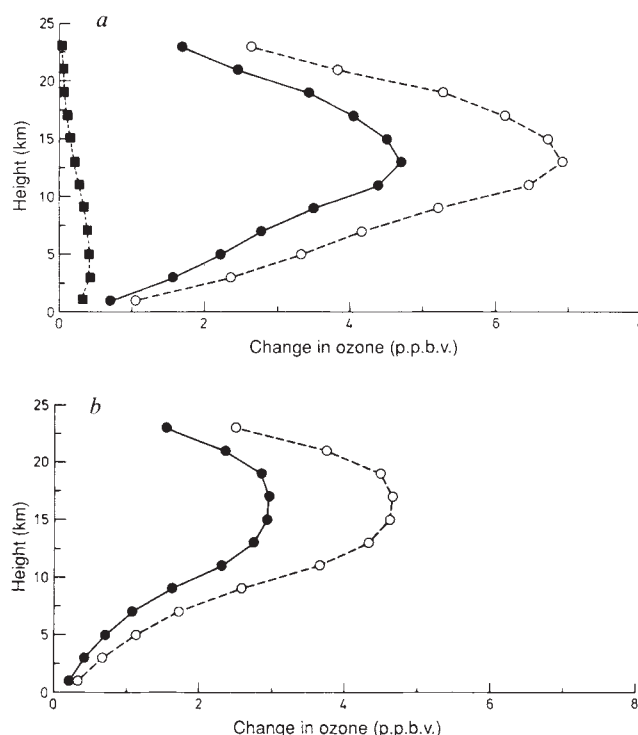


FIG. 2 Latitudinally averaged changes to the ozone mixing ratio profile from the baseline case caused by including estimated NO_x emission from aircraft and increasing the surface (man-made/industrial) emission source by 10%. (●) Aircraft emissions for 1990, (○) Aircraft emissions for 2000, (■) 10% increase to surface (man-made/industrial) NO_x emission. a, Northern Hemisphere; b, Southern Hemisphere. In the baseline case (A) the following global NO_x surface emissions¹⁹ were used (Tg yr^{-1} of NO_x): 69 (man-made/industrial), 16 (soils) and 26 (biomass burning), together with 26 from lightning distributed throughout the lowest 8 km of the model. In case B (1990, aircraft) we included an additional 1.94 Tg yr^{-1} of NO_2 ; case C (2000, aircraft) has an additional 3.18 Tg yr^{-1} , both compared with the baseline case (A). Case D (surface NO_x emission) has a further 10% of man-made or industrial emissions, that is 6.9 Tg yr^{-1} of NO_2 . The distribution of aircraft emissions for 1990 was assessed independently²¹. This was based on NO_x emission from each stage in the flight for six classes of aircraft. Information on flight frequency and distances of flights within and between five regions of the world were used to assess the vertical and latitudinal distribution of the emissions. The overall NO_x emission index produced from this study was $11.6 \text{ g (NO}_2\text{) per kg fuel}$, within the range given by a recent review²². No seasonal dependence in the emissions was assumed. Future aircraft emissions are difficult to assess, as they will depend on changes to engine technology as well as on air traffic growth and changes in their global distribution. For estimates of emission by 2000, a 5.1% annual growth in air traffic was assumed. This figure was selected from a range of projections, including those in ref. 23.

TABLE 1 Changes and sensitivities of model ozone and radiative forcing of surface temperature to NO_x emission changes

Case	Change from the baseline case (A)*				Sensitivity coefficients			
	ΔNO_2 emission (Tg yr ⁻¹ of NO ₂)	ΔO_3 inventory (Tg)	$\Delta T_0^\dagger$ (°C)	ΔOH inventory (Mg)	$\Delta\text{O}_3/\Delta\text{NO}_2$ (yr)	$\Delta T_0/\Delta\text{O}_3$ (°C Tg ⁻¹)	$\Delta T_0/\Delta\text{NO}_2$ (°C/(Tg yr ⁻¹))	$\Delta\text{OH}/\Delta\text{O}_3$
(B) Aircraft, 1990 (0–12 km)	1.94	15.2	0.014	7.3	7.8	9.2×10^{-4}	7.2×10^{-3}	0.48×10^{-6}
(C) Aircraft 2000 (0–12 km)	3.18	23.5	0.022	11.3	7.4	9.4×10^{-4}	6.9×10^{-3}	0.48×10^{-6}
(D) Surface emissions man-made	6.9	2.5	0.0017	1.8	0.36	6.7×10^{-4}	2.5×10^{-4}	0.71×10^{-6}

* See Fig. 2.

† Radiative forcing of surface temperature.

additional NO_x emissions are included. It is clear from the distribution of the ozone changes that the model sensitivity to aircraft NO_x emissions is greater in the Southern Hemisphere (which receives only 18% of the emissions) than in the Northern Hemisphere; the ozone content is ~60% more sensitive to emission changes in the Southern Hemisphere than it is to emission changes in the northern part of the Northern Hemisphere¹⁷. The sensitivity of ozone concentrations to surface NO_x emissions is clearly much less than to aircraft emissions and demonstrates the influence of nonlinear chemistry on ozone-production mechanisms^{24,25}. The absence of direct deposition to ground and slower conversion to nitric acid at lower temperatures lead to longer lifetimes for NO_x emitted from aircraft higher in the atmosphere¹⁷.

The distribution of the differences in ozone concentration between the baseline situation and the estimated 1990 case including aircraft emissions is shown in Fig. 3. The maximum changes of ~6 parts per 10⁹ (p.p.b.) occur at 12 km and are located at 50°N. There is a secondary peak in the Southern hemisphere of ~3 p.p.b. The maximum percentage changes are displaced to lower levels and towards the Equator. The ozone changes induced by surface NO_x emission are mainly confined to the Northern Hemisphere, lower in the atmosphere^{2,26}.

The height-dependent sensitivity of the radiative forcing of surface temperature to changes in the ozone profile⁸ was used to calculate changes to surface temperature from the data shown in Fig. 2, together with similar data for July. Table 1 shows the changes to the emission from the baseline case together with

predicted changes to the ozone inventory and to the surface temperature (ΔT_0) from the change in radiative forcing in our three model cases (1990 and 2000 aircraft emissions, and increased surface emissions). Also shown in Table 1 are the sensitivity coefficients for the change in model ozone inventory to changes in NO_x emissions, ($\Delta\text{O}_3/\Delta\text{NO}_2$), for the change ΔT_0 to changes in ozone, ($\Delta T_0/\Delta\text{O}_3$), and for the change ΔT_0 to change in NO_x ($\Delta T_0/\Delta\text{NO}_2$) emissions. It is clear that $\Delta\text{O}_3/\Delta\text{NO}_2$ is over 20 times greater for the cases of aircraft emissions than for surface emissions. The sensitivity coefficient $\Delta T_0/\Delta\text{O}_3$ is also greater for aircraft because of the ozone profile differences, and this causes further differences of a factor of 1.3. The final sensitivity coefficient $\Delta T_0/\Delta\text{NO}_2$ is therefore ~30 times greater for aircraft emissions than for surface emissions of NO_x. Aircraft may therefore contribute roughly as much to global warming as surface NO_x emissions from man-made sources, even though they contribute only ~3% to man-made NO_x emissions.

Table 1 also shows the changes ΔO_3 , ΔT_0 and sensitivity coefficients with increasing emissions from aircraft. The predicted change ΔT_0 of ~0.01 °C per decade for 1990–2000 can be compared with the change ΔT_0 from the observed CO₂ concentration increases over 1970–80, calculated in ref. 8 from the same radiative convective model as 0.067 °C.

An additional indirect effect of NO_x emissions is to increase tropospheric hydroxyl radical (OH) concentrations^{14,17,26}. Reaction with OH is the main removal processes of important greenhouse gases: methane and the hydrogen containing halocarbons such as HCFC-22 (ref. 1). Prediction of concentration increases of these gases therefore requires details on future changes to OH concentrations, and the indirect effect of reducing the tropospheric lifetimes of these gases will reduce the global warming potential (GWP) for NO_x emissions. The indirect effect of NO_x on the OH concentration is expected to depend on altitude because the main production route for OH is the reaction of O(¹D) atoms, formed from ozone photolysis, with water vapour (Fig. 1). Table 1 shows that the change in OH inventories relative to the change in O₃ inventory is ~40% lower for NO_x emitted from aircraft than for surface emission. This reduction may be largely attributed to the decrease in water vapour concentrations with altitude. Compared with surface NO_x emissions, the effect of aircraft NO_x emissions on methane and HCFCs will be further reduced because the resultant OH changes are largely confined to the region above 8 km (ref. 17) where the OH reaction rate coefficients are smaller because of the lower temperatures.

The sensitivity of the model mean tropospheric ozone concentration to surface NO_x emissions was calculated from case D in Fig. 2 to be 0.042 p.p.b. of ozone per Tg yr⁻¹ of NO₂. This is significantly lower than reported earlier¹⁰ for this model 0.19 p.p.b. of ozone per Tg yr⁻¹ of NO₂. The difference arises from an error in the calculation of the ozone inventory changes in the earlier work. As a consequence, the GWPs for the surface emission of NO_x in refs 1 and 10 are also in error. The corrected GWPs for surface-emitted NO_x are 30, 7, 2 (previously 150, 40,

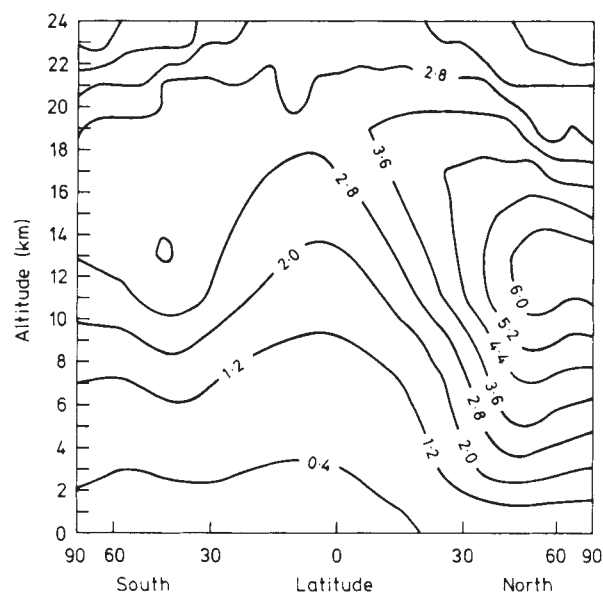


FIG. 3 Modelled changes in the distribution of the mixing ratio (p.p.b.v.) of tropospheric ozone caused by the inclusion of the 1990 estimate of aircraft NO_x emissions (January).

14) for integration to the time horizons of 20, 100 and 500 years, respectively, as expressed relative to carbon dioxide on the basis of a kilogram of emitted pollutant (CO_2 and NO_2). On the same basis the GWPs for NO_x emitted from aircraft, based on the data in Table 1, are 770, 210, 70 for time horizons of 20, 100 and 500 years. These figures have been calculated using the height-dependent sensitivity of ΔT_0 to ozone profile changes⁸. The global warming impact of surface-emitted NO_x has therefore been overestimated by a factor of around five in refs 1 and 10.

The 1990 estimates for surface man-made NO_x emission¹ of 66 Tg yr^{-1} (NO_2) and the value for aircraft emissions reported here of 1.9 Tg yr^{-1} (NO_2), together with the GWP figures above, give the relative global warming impact of each source. Combining the effects of both sources of NO_x , which are roughly equal, and using the man-made emissions source for CO_2 from ref. 1 we obtain a value of 0.035 for the relative global warming impact of NO_x compared with CO_2 at the 100-year time horizon. $\square$

Received 16 July; accepted 10 October 1991.

- Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. (eds) *Climate Change: the IPCC Scientific Assessment* (Cambridge University Press, 1990).
- Hough, A. M. & Derwent, R. G. *Nature* **344**, 645–650 (1990).
- Thompson, A. M., Huntley, M. A. & Stewart, R. W. *J. geophys. Res.* **95**, 9829–9844 (1990).
- Logan, J. A. *J. geophys. Res.* **90**, 10463–10482 (1985).
- Volz, A. & Kley, D. *Nature* **332**, 240–242 (1988).
- Bojkov, R. D. *Tropospheric Ozone Regional and Global Scale Interactions*, NATO ASI Series C, Vol. 227 (ed. Isaksen, I.S.A.) 83–96 (NATO, 1988).
- Ramanathan, V., Singh, H. B., Cicerone, R. J. & Kiehl, J. T. *J. geophys. Res.* **90**, 5547–5566 (1985).
- Lacis, A. A., Wuebbles, D. J. & Logan, J. A. *J. geophys. Res.* **95**, 9971–9981 (1990).
- Wang, W. C. & Sze, N. D. *Nature* **286**, 589–590 (1980).
- Derwent, R. G. *Trace Gases and their Relative Contribution to the Greenhouse Effect AERE R13716* (HMSO, London, 1990).
- Hildago, H. & Crutzen, P. J. *J. geophys. Res.* **82**, 5833–5866 (1977).
- Isaksen, I. S. A. *Proc. Quadrenn. Ozone Symp.*, Vol. II (ed. London, J.) 845–852 (NCAR, Boulder, 1980).
- Derwent, R. G. *Atmos. Envir.* **16**, 1997–2007 (1982).
- Beck, J. P. et al. *The Effect of Aircraft Emissions on Tropospheric Ozone in the Northern Hemisphere*, *Atmos. Envir.* **A26**, 17–19 (1992).
- Crutzen, P. J. & Bruhl, C. *Air Traffic and the Environment*, Lecture Notes in Engineering, Vol. 60 (ed. Schumann, U.) 96–106 (Springer, Berlin, 1990).

- Wuebbles, D. J. & Kinnison, D. E. *Air Traffic and the Environment*, Lecture Notes in Engineering, Vol. 60 (ed. Schumann, U.) 107–123 (Springer, Berlin, 1990).
- Johnson, C. E. & Henshaw, J. *The Impact of NO_x Emissions from Tropospheric Aircraft*, AEA-EE-0127 (Harwell Laboratory, 1991).
- Johnston, H. W., Kinnison, D. E. & Wuebbles, D. J. *J. geophys. Res.* **94**, 16351–16363 (1989).
- Hough, A. M. *J. geophys. Res.* **96**, 7325–7362 (1991).
- Hough, A. M. *Atmos. Envir.* **23**, 1235–1261 (1989).
- McInnes, G. & Walker, C. T. *The Global Distribution of Aircraft Air Pollutant Emissions Rep. LR 872* (AP) (Warren Spring Laboratory, in the press).
- Egli, R. A. *Chimia* **44**, 369–371 (1990).
- Nüßer, H.-G. & Schmitt, A. *Air Traffic and the Environment*, Lecture Notes in Engineering, Vol. 60 (ed. Schumann, U.) 1–11 (Springer, Berlin, 1990).
- Lin, X., Trainer, M. & Liu, S. C. *J. geophys. Res.* **93**, 15879–15888 (1988).
- Liu, S. C. et al. *J. geophys. Res.* **92**, 4191–4207 (1987).
- Hough, A. M. & Johnson, C. E. *Atmos. Envir.* **A25**, 1819–1835 (1991).

ACKNOWLEDGEMENTS. We thank R. G. Derwent for stimulating this research, L. Michaelis discussions, and R. G. Derwent, D. Warrilow, M. J. Norgett and D. R. Woodwork for comments. This work was funded by DGXII of the Commission of the European Communities. Model development was funded by the UK Department of the Environment.

Changes in frequency–size relationship from small to large earthquakes

Javier F. Pacheco, Christopher H. Scholz & Lynn R. Sykes

Lamont-Doherty Geological Observatory and Department of Geological Sciences, Columbia University, Palisades, New York 10964, USA

THE constant 'b value' observed in frequency–magnitude distributions of earthquakes has been taken as an indication of self-similarity at all magnitudes. Hence, earthquake properties should scale uniformly in the same way from small to large earthquakes. It has often been observed, however, that the seismic moment released in small earthquakes scales differently with rupture length than it does for large events^{1,2}, where the crossover between small and large events is defined at a rupture dimension equal to the down-dip width of the seismogenic zone. A possible explanation for this contradiction is that frequency–size distributions are biased by small earthquakes. Small events dominate most global earthquake catalogues because of the short time-period covered. Another source of bias in size distributions is the saturation of earthquake magnitudes for large events. Here we correct biases in calculations of b values and present evidence for a change in b value in frequency–size distributions. We find that a break in self-similarity, from small to large earthquakes, occurs at a point where the dimension of the event equals the down-dip width of the seismogenic layer.

The observed rate of earthquakes ($\dot{n}$) with seismic moment $\geq M_0$ or energy $\geq E$ follows a power-law distribution: $\dot{n}(M_0) = aM_0^{-b}$. This observation was first made by Gutenberg and Richter³ who expressed it in terms of earthquake magnitude, M , as $\log_{10}[\dot{n}(M)] = A - bM$. The parameter A measures the level of seismic activity in the region. For small earthquakes the slope (b value) is found to lie between 0.8 and 1.2. A constant b value for the entire range of magnitudes implies self-similarity of earthquakes, that is, earthquake properties scale in the same way from small to large events.

Rydelek and Sacks⁴ found a significant deviation from self-similar behaviour for earthquakes with magnitudes smaller than 2.0. As geometrical constraints are different for small and large earthquakes¹, we also expect to find another departure from self-similarity at larger magnitudes. Small earthquakes can grow in both length (L) and width (W); their rupture dimensions have no bounds. Large earthquakes, in principle, have no bounds in rupture length, but their down-dip width is limited by the thickness of the seismogenic layer (the region capable of generating earthquakes). They are confined between the Earth's surface and the brittle–ductile transition at depth (Fig. 1). Thus, small earthquakes are confined by the two dimensions of the fault but large ones by only one dimension. This difference in dimensionality has consequences for scaling properties from small to large earthquakes. Seismic moment scales as L^3 for small events and as L^2 for large ones^{1,2}. Source spectra for acceleration in the far field increase at intermediate frequencies as ω^2 for small earthquakes and as $\omega^{5/4}$ for large ones⁵.

Rundle⁶ derived frequency–magnitude relations for small and large earthquakes. He shows, with a tiling argument, that to preserve scale invariance under a change of spatial scale, the b value has to change from 1.0 to 1.5, from small to large earthquakes. Therefore, a break in self-similarity should be observable as a change in the slope of a frequency–size diagram.

In Fig. 2 we plot a frequency–magnitude distribution using a worldwide catalogue⁷ of shallow earthquakes (depth ≤ 60 km)

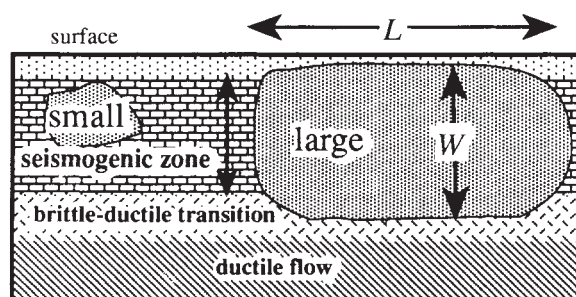


FIG. 1 Two types of earthquakes; small (unbounded) and large (bounded). L is rupture length, along strike of fault, W down-dip width of the rupture.

that covers the years 1900 to 1989 and is complete for $M_s \geq 7.0$. In Fig. 2a we use the surface-wave magnitude M_s of Abe⁸ and in Fig. 2b the moment-magnitude (M_w). Visual inspection of the figures shows pronounced changes in the slope at magnitudes M_s (Abe) = 7.5 and M_w = 7.5, respectively. We fitted two lines through the data (solid lines) using a robust M -estimate with a maximum likelihood formula⁹. We applied a cutoff at magnitude 8.0 to avoid saturation of 20-s surface-wave magnitudes¹⁰ when modelling the data from Fig. 2a. For the data shown in Fig. 2b, we applied a cutoff at magnitude M_w = 8.8 as the interval of 90 years seems to be too short to sample adequately events of larger magnitude. To find the change in b value, we maximize the difference between slopes given by a Student's t test for unequal variance and minimize the standard deviations of the best-fit linear parameters. Doubling the number of parameters, from one to two straight lines, is not the main cause of the variance reduction; the best fits are achieved with two lines, as is revealed by the Akaike Information Criterion^{2,11} (Table 1). Other global catalogues, such as the Preliminary Determination of Epicenters and the Bulletin of the International Seismological Centre, also show changes in the b value at similar magnitudes. These catalogues, however, do not cover more than 25 years in M_s determinations and do not uniformly assign surface-wave magnitudes in time. The Harvard Centroid Moment Tensor Solutions (CMTS) catalogue reports a uniform magnitude based on the seismic moment for global events with magnitudes $M_w \geq 5.0$ from 1977 through the present, and is complete⁷ at $M_w \geq 5.6$. The break in self-similarity occurs at M_w = 7.4 (Fig. 3a). The change in slope is not as well defined because the period covered by the catalogue (14 years) is too short to provide adequate statistics for earthquakes with $M_w \geq 7.8$. But the distribution for magnitudes between 5.6 and 7.5 follows a power law with a b value of 1.07, in broad agreement with the b value for Fig. 2 for magnitudes between 7.0 and 7.5.

Global catalogues of shallow earthquakes with magnitude larger than 7.0 are dominated by seismicity at subduction zones⁷. The observed break in b value in a worldwide catalogue must be related to the dimensions of the seismogenic zone for subduction zones, that is, the down-dip extent of earthquake activity along the plate boundary. The seismic moment and the area of rupture in an earthquake (S) are related by $M_0 = \mu u S$, where μ is the shear modulus, u is the average displacement and¹² $S = LW = 5(\mu u / \Delta\sigma)^2$. Assuming a stress drop¹³ ($\Delta\sigma$) of 3 MPa, and a magnitude M_w between 7.5 and 7.6, the down-dip width of the seismogenic zone corresponding to these values is ~55–62 km. It has been observed^{14,15} that the down-dip width of the thrust interface at subduction zones varies between 50 and

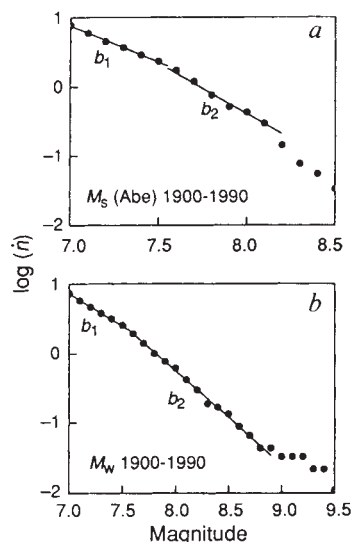


FIG. 2 Frequency-size distribution using Pacheco-Sykes⁷ catalogue. Frequency of earthquakes as a function of magnitude: a, M_s (Abe); b, moment-magnitude M_w . n is rate of occurrence of events per year of a given magnitude or larger. Subscripts to b denote segments fitted by linear relationships of Table 1.

Figure	tude break	A1	A2	b1	b2	Reduction in Akaike number (%)
2a	7.5	8.1 ± 0.2	11.7 ± 0.6	1.04 ± 0.03	1.51 ± 0.08	44
b	7.5	7.2 ± 0.2	10.3 ± 0.3	0.90 ± 0.02	1.31 ± 0.03	47
3a	7.4	8.4 ± 0.1	12.2 ± 0.9	1.07 ± 0.01	1.6 ± 0.1	61
b	6.4	6.6 ± 0.1	9.4 ± 0.3	0.90 ± 0.01	1.34 ± 0.05	71
c	6.4	6.4 ± 0.1	8.5 ± 0.2	0.92 ± 0.01	1.24 ± 0.03	67
d	5.9	6.4 ± 0.2	9.4 ± 0.3	0.88 ± 0.04	1.39 ± 0.04	26

200 km. As smaller events and interface widths are much more frequent, we expect the break in slope to be associated with the smaller values of W .

In Fig. 3 we compare the frequency distributions of intermediate and deep earthquakes from the CMTS catalogue. The break in b value for intermediate (100 km ≤ depth ≤ 300 km) and deep (depth > 300 km) events is at a magnitude between 6.4 and 6.5 (Fig. 3b and c), suggesting a width of the seismogenic zone from 5 to 30 km, for a stress drop¹⁶ between 1 and 100 MPa. For shallow transform faults (Fig. 3d) the change in slope occurs between magnitude M_w = 5.9 and 6.0, for a width of ~10 km. This catalogue for transform faults does not seem to be complete for magnitudes smaller than 5.8. Thus, the slope of 0.88 obtained for small earthquakes may not reflect the actual b value for transform faults. For the southern California region, a transform-fault area, however, we find a slope of 0.94 for events of magnitude 3.0 to 6.0 for the period 1935 to 1985, as compared with a slope of 1.39 for $M_w \geq 6.0$ in Fig. 3d (Table 1). The former data are taken from the catalogue of the Caltech-southern California seismic network, which appears to be complete at the magnitude 3.0 level since 1932. Along transform faults and fracture zones the width of the seismogenic zone is in the range 5–15 km (ref. 17). The down-dip width of the faults that break during intermediate and deep earthquakes is not well known; however, we expect it to be between that for transform faults and subduction zone earthquakes along the thrust plate interface.

For more than 50 years it was thought that earthquake size distribution extended over all magnitude ranges with a b value

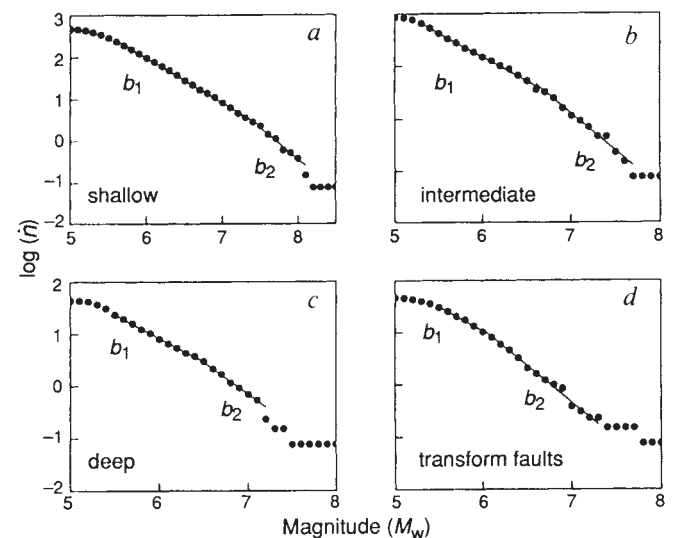


FIG. 3 Frequency-size distribution of worldwide earthquakes in CMTS catalog from 1977 to 1990. Depth of events: a, shallow earthquakes (≤ 70 km); b, intermediate (100–350 km); c, deep earthquakes (≥ 350 km); d, only shallow earthquakes along transform faults. n is rate of occurrence of events per year of a given magnitude or larger. Subscripts to b denote segments fitted by linear relationships of Table 1.

universally close to 1. Although Gutenberg and Richter³ recognized a curvature in the distribution for large magnitudes, this was thought to be due either to incompleteness of the early catalogue or, later, to magnitude saturation¹⁸. Having overcome these difficulties, we now see that there is a break in the slope. The b values of ~ 1 and 1.5 agree with Rundle's⁶ predictions.

That large and small earthquakes have different size distributions has important implications for earthquake hazard analysis, which often relies on predicting the former from extrapolation of the later. For an individual fault, we know that this difference results in large earthquakes occurring more frequently than expected from small earthquake activity^{19,20}. Here we have found, conversely, that for a system containing many faults, far fewer large events will occur than expected from extrapolating the distribution of small earthquakes. □

Received 5 July; accepted 14 October 1991.

1. Scholz, C. H. *Bull. Seism. Soc. Am.* **72**, 1–14 (1982).
2. Shimazaki, K. in *Earthquake Source Mechanics*. AGU Geophys. Monogr. **37** (eds Das, S., Boatwright, J. & Scholz, C.) 209–216 (American Geophysical Union, Washington, DC, 1986).
3. Gutenberg, B. & Richter, C. *Seismicity of the Earth and Associated Phenomena*, 2nd edn, 310 (Princeton University Press, 1956).
4. Rydelek, P. A. & Sacks, I. S. *Nature* **337**, 251–253 (1989).
5. Boatwright, J. & Choy, G. *J. geophys. Res.* **94**, 15541–15553 (1989).
6. Rundle, J. B. *J. geophys. Res.* **94**, 12337–12342 (1989).
7. Pacheco, J. F. & Sykes, L. R. *Bull. Seism. Soc. Am.* (submitted).
8. Abe, K. *Phys. Earth planet. Inter.* **27**, 72–92 (1981).
9. Press, W., Flannery, B., Teukolsky, S. & Vetterling, W. *Numerical Recipes in C: The Art of Scientific Computing*, 735 (Cambridge University Press, New York, 1988).
10. Geller, R. *Bull. Seism. Soc. Am.* **66**, 1501–1523 (1976).
11. Akaike, H. *IEEE Trans. Autom. Control* **AC-19**, 716–723 (1974).
12. Rundle, J. B. & Kanamori, H. *J. geophys. Res.* **92**, 2606–2616 (1987).
13. Kanamori, H. & Anderson, D. *Bull. Seism. Soc. Am.* **65**, 1073–1095 (1975).
14. Sykes, L. R. *J. geophys. Res.* **76**, 8021–8041 (1971).
15. Sykes, L. R. & Quittmeyer, R. C. *Earthquake Prediction: An International Review*, Maurice Ewing Ser. 4 (eds Simpson, D. & Richards, P.) 217–247 (American Geophysical Union, Washington DC, 1981).
16. Frohlich, C. *Ann. Rev. Earth planet. Sci.* **17**, 227–254 (1989).
17. Marone, C. & Scholz, C. *Geophys. Res. Lett.* **15**, 621–624 (1988).
18. Chinnery, M. A. & North, R. G. *Science* **190**, 1197–1198 (1975).
19. Davison, F. & Scholz, C. *Bull. Seism. Soc. Am.* **75**, 1349–1362 (1985).
20. Scholz, C. H. in *Spontaneous Formation of Space-Time Structures and Criticality* (eds Riste, T. & Sherrington, D.) 41–56 (Kluwer, Dordrecht, 1991).

ACKNOWLEDGEMENTS. We thank K. Jacob, A. Malinverno and C. Kisslinger for reviewing the manuscript and suggesting improvements. This work was supported by the NSF and SCEC.

Food-web analysis through field measurement of per capita interaction strength

R. T. Paine

Department of Zoology, University of Washington, Seattle, Washington 98195, USA

THE idea that the connections between species in ecological assemblages are characterized by a trait called 'interaction strength' has become a cornerstone of modern ecology. Since the classic paper of Watt¹, ecologists have acknowledged the importance of distinguishing pattern (static community features) from process (dynamically based mechanisms), which can determine the immediate observed details. Food webs display pathways of implied dynamics, and thus potentially unite pattern and process in a single framework^{2,3}. Most analyses^{4–8} have focused entirely on web topology and the derived descriptive properties. By contrast, attempts to generalize how natural communities are organized^{9,10} or summary statements about whole communities^{6,11–13} have emphasized critical processes. Elton's² insights and May's³ generalizations and analyses have stimulated current developments. In May's approach, the idea of interaction strength is precise, reflecting coefficients in a jacobian matrix associated with a community dynamics model. He found striking dependence of community stability both on web complexity and on the number and

strength of interactions. By contrast, empiricists have usually determined relative interaction strength from single-species removals analysed by multivariate statistics^{14,15}. I report here the first experimental study designed to estimate interaction strengths in a species-rich herbivore guild, documenting on a per capita basis mainly weak or positive interactions, and a few strong interactions, a pattern which has profound implications for community dynamics.

I sought a procedure for quantifying the magnitude of a consumer's influence on an entire assemblage, thus including direct and potentially indirect effects, that would be based on manipulation of single-species densities and that would permit direct comparison against some common standard or baseline condition. The most important interactions in a community context are those changing the density of a species which, in the absence of natural enemies or disturbance, would come to dominate the assemblage. Many marine examples exist in which monocultures formed by the local competitive dominant develop on shallow water, hard surfaces^{16,17}. The potential formation of a monoculture provides a common baseline and thus the means of calibrating the impact of individual species on an assemblage. A species capable of preventing the development of a monoculture, or destroying one already established, should be a strong interactor in the sense that its density and activities should have measurable consequences for the competitively superior prey, and thus indirectly for other members of the local assemblage. A consumer characterized by minimal effect on the dominant is apt to be of little direct significance to the whole assemblage, although its activities might influence some other prey. It would thus be a weakly interacting species. Here I describe a direct evaluation of interaction strength, one which permits unambiguous quantification and comparison. The derived metric is based on an estimate of per capita effect, distinguishes between biologically important in addition to statistically significant

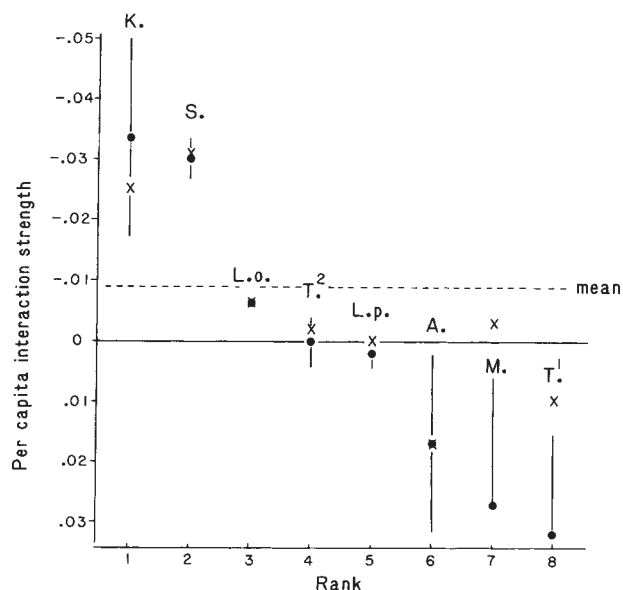


FIG. 1 Per capita interaction strength, measured as a grazer's capacity to influence the establishment of brown algal sporelings, ranked from negative to positive values. The bootstrap procedure calculated the index $E-C/Cd$, where d represents the density of the specific replicate, for as many times as that treatment was replicated. From these data a mean and s.e. were calculated. This procedure was then repeated 100 times; the means of these means and standard errors are given in Table 1 and graphed here (●, bootstrap means and their s.e.). The procedure mirrors the field experimental protocol and preserves the sensitivity of s.e. estimates to the number of replicates. Field experimental means (×) are also shown. S., *S. purpuratus*; K., *K. tunicata*; L.o., *L. ochraceus*; M., *M. hindsii*; L.p., *L. painei*; T., *T. lineata*; A., *A. mitra*. Exact values and sample details are given in Table 1. The dashed line is the mean of all species values, and is negative as expected in a predator-prey relationship.

TABLE 1 Measured and derived data on experimentally excluded grazers

Taxon/species	Treatment replicates	Natural density (number m ⁻²)	Range of experimental density (number m ⁻²)	Experimental per capita effect	Effect on % sporeling change	Bootstrapped per capita effect (±1 s.e.)
Sea urchin						
<i>Strongylocentrotus purpuratus</i>	5	25.4	15–44	–0.031	–78.7	–0.030(±0.003)
Chiton						
<i>Katharina tunicata</i>	4	28.0	8–30	–0.025	–70.0	–0.033(±0.016)
<i>Mopalia hindsii</i>	3	1.9	20–42	–0.003	–0.6	+0.027(±0.022)
<i>Tonicella lineata</i> /first series	6	12.3	30–83	+0.005	+6.2	+0.033(±0.022)
<i>Tonicella lineata</i> /second series	8	12.3	36–500	–0.002	–2.5	+0.000(±0.004)
Limpet						
<i>Acmaea mitra</i>	7	4.3	15–22	+0.017	+7.3	+0.017(±0.016)
<i>Lottia painei</i>	9	3.2	159–271	–0.001	–0.3	+0.002(±0.002)
<i>L. ochraceus</i>	1	0.1	115	–0.007	–0.1	–0.006
Grazer-free controls	41					

Total grazer enclosures, or exclosures, for individual species were made from rings or 'doughnut' of Sea Goin' Poxy PuttyTM surrounding natural rock surfaces at about the 0.0 m tidal level. Once the putty had hardened, it was dried and painted with copper-based antifouling paint and all resident brown algae were removed from the surrounded surface. Whenever possible the area of the experimental arena was scaled to the needs and natural population density of the target species. This was not always possible, and thus for the smaller bodied or less common species, treatment densities tend to greatly exceed natural densities. The product of experimental per capita effect × natural density suggests the possible influence of each species on the early survival of a complex of brown algal sporelings, the reference state against which all impacts were judged. Details of the bootstrap procedure used to estimate mean (and s.e. of that mean) per capita effect are provided in the text and Fig. 1.

influences, and thus relates immediately to the important applied matter of whether a single species has the capacity to alter the structure of an entire assemblage.

Experiments were conducted at Tatoosh Island (48°24'N, 124°44'W) on the exposed outer coast of Washington State, United States. There, as at many comparable sites, the mid-low intertidal zones are dominated by stands of laminarial brown algae¹⁸; traditional schemes of description have even characterized these low intertidal zones by their dominant species¹⁹. At Tatoosh the canopy-forming algae are *Hedophyllum sessile* in the presence of grazers¹⁸, and *Alaria* spp. when grazers are absent or experimentally excluded¹⁷. Under this canopy, space tends to be dominated by 10–15 species of coralline and numerous fleshy red algae. A phylogenetically diverse guild of grazers consume these plants, including crabs, sea urchins, chitons, limpets, a few fishes and an ecologically unknown set of mesograzers²⁰, mainly amphipods. These results concern only a subset of these grazers, but I believe they incorporate all but one of the ecologically important species. Treatments and number of replicates are listed in Table 1. The experiments were generally initiated in mid-August and terminated the following April. Control treatments ($N=41$) were not invaded by macrograzers, and suggest that elevated, painted surfaces were effective enclosures/exclosures for molluscan grazers. Similarly, the paint has little effect on the surrounding algae (L. Johnson, personal communication). To constrain sea urchins a wire barrier enclosed the painted surface. I used the density of brown algal sporelings <5 cm long that had appeared by April in control (grazer-free) treatments as the common reference state. Counts of sporeling numbers suggest domination by *Alaria* (March 1990, 93% of 1,026 plants; April 1990, 92% of 469 plants). The change relative to the mean control value was then determined for each replicate on a m² basis, and expressed as the ratio:

$$\frac{\text{Treatment density} - \text{Control density}}{\text{Control density}} \quad \text{or} \quad \frac{E - C}{C}$$

The ratio varies from –1.0 (total annihilation of sporelings) to 0 (no apparent effect) to some theoretically unlimited positive value in cases where brown algal recruitment or survival was enhanced. This value was then adjusted to a per capita basis by dividing by the grazer density characteristic of that specific replicate. Because the index embodies average values for control yet independently estimated values for grazer and algal density,

the full range of variation inherent in the experimental protocol is not adequately expressed by parametric statistical procedure. Instead, the entire data set was bootstrapped²¹ to recreate the field experiment so that there were paired experimentals and controls. Mean per capita effect on the basis of field data and means and standard errors on the basis of the bootstrap are given in Table 1 and Fig. 1.

Per capita influence on sporeling density showed both negative and positive values, with an average impact of –0.009 (or –0.9% m⁻² per 7.5 months). A net negative value should characterize a guild of consumers. The positive effects are explicable because these species consume crustose coralline algae, which in turn can inhibit algal spore recruitment^{22,23}. Although this data set does not include all members of the grazer guild, it is clear that the majority of the species (5 of 7) are relatively benign, and either have no significant negative influence on the

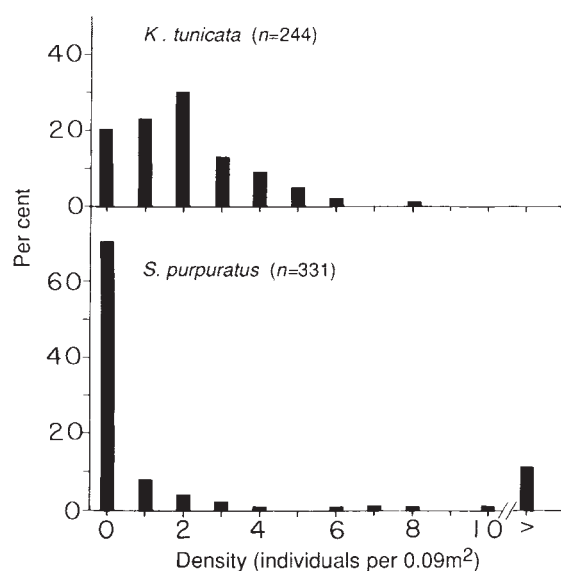


FIG. 2 Density estimates of the two invertebrate grazers characterized by strong trophic interactions at Tatoosh Island. Quadrat dimensions were 30 × 30 cm; n , number of samples taken 1988–1991. The chiton *Katharina* density is adequately described by a fit to a Poisson series ($\chi^2=13.7$, $P>0.05$). Urchin densities are extremely under-dispersed, indicating extensive aggregation.

reference state (*Lottia painei*; *Tonicella*, second series) or interact positively (*Acmaea mitra*; *Mopalia*; *Tonicella*, first series). The first series of *Tonicella* experiments, conducted at relatively low algal recruitment densities (510 m^{-2}), suggests recruitment enhancement, presumably because invadable substratum is generated by the grazers' foraging. At the much higher algal densities ($4,800\text{ m}^{-2}$) characteristic of the second series, the influence of this weakly rasping chiton was swamped, and no interaction effect is detectable.

The two strongly negative species (*Katharina* and *Strongylocentrotus purpuratus*) are statistically indistinguishable (analysis of variance; $F=0.41$, $P>0.5$). Their effects are not additive because urchins at densities greater than about 30 m^{-2} exclude *Katharina* (by eating them or driving them away). Hence these strong interactors do not coexist at their average (Table 1) densities although sparsely mixed populations are commonplace. Sea urchins in particular are widely acknowledged to suppress development of brown algal canopies, in the process producing 'barrens'²³. Calculations based on average Tatoosh densities in the low intertidal (25.4 m^{-2}) and the experimentally estimated per capita effect (-0.031) predict that some brown algae should be able to recruit successfully. But urchins are not haphazardly distributed (Fig. 2), existing in aggregations in which local density often exceeds 200 m^{-2} . Under such conditions, no fleshy algae survive, implying that factors spatially concentrating a strong individual effect will make a significant contribution to community organization, as in insects^{24,25}. By contrast, *Katharina* is randomly or haphazardly distributed (Fig. 2). Calculation of local effect indicates that about 70% of the algae ($28\text{ m}^{-2} \times -0.025$) will be consumed during the sporeling stage. The surviving density is more than sufficient to produce a 100% brown algal canopy, even in the presence of numerous weakly interacting grazers.

I chose to work with a guild of invertebrates generally considered to be 'grazers'. The presence of interactions of both signs significantly departing from zero (no effect) (Fig. 1) in closely related taxa (for example in both chitons and limpets) warns against the assumption of simple additivity, that is, that all guild members will have the same sign to their impact on the community, and implies that compensatory interactions could be commonplace. Equally, the exaggerated density of some grazers (Table 1) suggests that the effects are probably real. If a negative effect cannot be detected at 2–900 times the natural density, there are probably no circumstances under which it might exist.

I have thus presented an approach to quantifying per capita interaction strength in ecological communities that is practical, that allows statistical inference while distinguishing biological importance, and that summarizes and even predicts the capacity of a single species to alter the structure of an entire assemblage. When applied to a rocky intertidal community, 5 of 7 consumer species interact weakly or even positively with a competitively dominant prey. If communities are generally typified by such patterns of a few strong interactions embedded in a majority of negligible effects, the implications for community stability and organization are substantial. Significantly, none of this information can be obtained without experimental manipulation, regardless of how convenient or appealing it might be to infer the intensity and consequences of interspecific interactions from observation alone. Alternative but promising experimental approaches²⁶ to identifying whether particular predator–prey interactions are strong or weak, in the context of community-wide influence, have yet to proceed beyond single-species evaluations. The strong skew of interaction strengths (Table 1) bears important implications for the management and conservation of natural assemblages. When combined with estimates of density and tendencies toward aggregation, the relative importance and potential involvement of a species in chains of indirect effects can be identified. Management, conservation and even reconstruction of natural assemblages will require such information. □

Received 15 July; accepted 4 October 1991.

1. Watt, A. S. *J. Ecol.* **35**, 1–22 (1947).
2. Elton, C. *Animal Ecology* (Sidgwick and Jackson, London, 1927).
3. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton University Press, 1973).
4. Cohen, J. E., Luczak, T., Newman, C. M. & Zhou, Z. M. *Proc. R. Soc. B* **240**, 607–627 (1990).
5. Pimm, S. L. *Food Webs* (Chapman and Hall, London, 1982).
6. Paine, R. T. *J. anim. Ecol.* **49**, 667–685 (1980).
7. Lawton, J. H. in *Ecological Concepts* (ed. Cherrett, J. M.) 43–78 (Blackwell Scientific, Oxford, 1989).
8. Pimm, S. L., Lawton, J. H. & Cohen, J. E. *Nature* **350**, 669–674 (1991).
9. Hairston, N. G., Smith, F. E. & Slobodkin, L. B. *Am. Nat.* **94**, 421–425 (1960).
10. Menge, B. A. & Sutherland, J. P. *Am. Nat.* **130**, 730–757 (1987).
11. Carpenter, S. R., Kitchell, J. F. & Hodgson, J. R. *Bioscience* **35**, 634–639 (1985).
12. Power, M. E. *Science* **250**, 811–814 (1990).
13. Fretwell, S. D. *Oikos* **50**, 291–301 (1987).
14. Fowler, N. J. *Ecol.* **69**, 843–854 (1981).
15. Menge, B. A. & Farrell, T. M. *Adv. Ecol. Res.* **19**, 189–262 (1989).
16. Lubchenco, J. *Am. Nat.* **112**, 23–39 (1978).
17. Paine, R. T. *Ecology* **65**, 1339–1357 (1984).
18. Dayton, P. K. *Ecol. Monogr.* **41**, 351–389 (1971).
19. Stephenson, T. A. & Stephenson, A. *A Life Between Tidemarks on Rocky Shores* (Freeman, San Francisco, 1972).
20. Brawley, S. H. & Fei, X. G. *J. Phycol.* **23**, 614–623 (1987).
21. Efron, B. & Tibshirani, R. *Science* **253**, 390–395 (1991).
22. Breitburg, D. L. *Ecology* **65**, 1136–1143 (1984).
23. Johnson, C. R. & Mann, K. H. *J. exp. mar. Biol. Ecol.* **96**, 127–146 (1986).
24. Hassell, M. P. *The Dynamics of Arthropod Predator–Prey Systems* (Princeton University Press, 1978).
25. Kareiva, P. *Nature* **326**, 388–390 (1987).
26. Juliano, S. A. & Lawton, J. H. *Oecologia* **83**, 535–540 (1990).

ACKNOWLEDGEMENTS. I thank J. T. Wootton, C. Pfister, M. E. Power and J. Ruesink for criticism, G. Gilchrist and T. Ives for statistical advice and P. Kareiva for scepticism. My research on Tatoosh is possible only with permission from the Makah Tribal Council and the US Coast Guard and is supported by the NSF.

Peripheral nerve injury triggers central sprouting of myelinated afferents

Clifford J. Woolf, Peter Shortland & Richard E. Coggeshall*

Department of Anatomy and Developmental Biology, University College London, Gower Street, London WC1E 6BT, UK

* Department of Anatomy and Neuroscience, University of Texas Medical Branch, University Blvd, Galveston, Texas 77550, USA

THE central terminals of primary afferent neurons are topographically highly ordered in the spinal cord¹. Peripheral receptor sensitivity is reflected by dorsal horn laminar location: low-threshold mechanoreceptors terminate in laminae III and IV (refs 2, 3) and high-threshold nociceptors in laminae I, II and V (refs 4, 5). Unmyelinated C fibres, most of which are nociceptors⁶, terminate predominantly in lamina II (refs 5, 7). There is therefore an anatomical framework for the transfer of specific inputs to localized subsets of dorsal horn neurons. This specificity must contribute to the relationship between a low-intensity stimulus and an innocuous sensation and a noxious stimulus and pain. We now show that after peripheral nerve injury the central terminals of axotomized myelinated afferents, including the large A β fibres, sprout into lamina II. This structural reorganization in the adult central nervous system may contribute to the development of the pain mediated by A-fibres that can follow nerve lesions in humans^{8,9}.

Peripheral nerve injury in humans is associated with decreased or abnormal sensations, including pain^{8–12}. Here we examined whether structural changes in the central terminals of the axotomized primary afferent neurons have a particular role in the development of these sensory disorders. Our rationale for this is the finding that after peripheral nerve injury the expression of the developmentally regulated phosphoprotein GAP-43, which is associated with or is a marker of axonal growth¹³, is increased in small and large dorsal root ganglion cells^{14–16} and is distributed not only to the damaged axonal tip in the

periphery¹⁷, but also to the central terminals of the axotomized afferents in the spinal cord and brain stem¹⁸. GAP-43 is also present in the dorsal horn in growth cone-like structures¹⁹. Peripheral nerve injury also results in a substantial reduction in the synaptic contacts made by primary afferents with neurons in lamina II (ref. 20), presumably by axonal withdrawal or some other form of transganglionic²¹ or transsynaptic degenerative change²². The combination of vacant synaptic sites in lamina II and the presence in axon terminals of growth-associated proteins may provide the circumstances for sprouting of axonal terminals to this area. Vacant synapses alone may not be enough to induce sprouting, explaining the relative failure of collateral sprouting of intact afferent central terminals into denervated regions of the adult spinal cord²³. We have examined the sprouting of the central terminals of axotomized afferents using two approaches: bulk labelling of myelinated afferents and single afferent labelling.

Application of the neural tracer horseradish peroxidase (HRP) to a peripheral nerve results in its transganglionic transport to the central terminals of the labelled axons²⁴. This approach can be refined using lectins that are selectively transported by particular subgroups of afferents. Wheatgerm agglutinin, when conjugated to HRP, predominantly labels C-fibre terminals in lamina I and II (refs 25, 26) (Fig. 1a). By contrast, the B unit of cholera toxin (cholera toxinogenoid) (CB) when conjugated to HRP only labels the central terminals of myelinated afferents²⁷. Cholera toxinogenoid binds to the GM1 ganglioside

receptor that is present almost exclusively on those dorsal root ganglion cells that possess myelinated axons^{28,29}.

In the eight animals in which we applied CB-HRP to intact nerves (sciatic 3, sural 5), label in the dorsal horn was only found in laminae I, III and deeper, compatible with A, but not C-fibre central terminal labelling (Fig. 1b, c). One week after a sural nerve section and ligation, the distribution of CB-HRP label changed with a more dorsal extension of the label from lamina III into lamina II (Fig. 1d). By two weeks postsection, the labelled sural central terminals extended throughout lamina II forming a dense continuous band from lamina I to V (Fig. 1e). This changed central terminal pattern was present when sural nerves were labelled from 2 to 15 weeks postsection ($n = 8$; Fig. 1f, g, h) and was restricted to the sural nerve territory. This was particularly noticeable when the entire sciatic nerve was labelled (Fig. 1f, g, h). The extension of label into lamina II was not a consequence of unmyelinated afferents developing a CB binding capacity. This was shown by the distribution of CB-HRP label in axonal profiles of the L4 dorsal root after chronic sciatic nerve section. Of 141 axons counted, 29 of 61 myelinated axons were labelled but only 5 out of 80 unmyelinated axons. These numbers are very similar to the distribution of CB labelling to cells with myelinated and unmyelinated axons in the intact rat dorsal root ganglion²⁸. These observations do not exclude the possibility that C-fibre terminals also sprout after peripheral axotomy.

The redistribution of A-fibre terminals into lamina II could

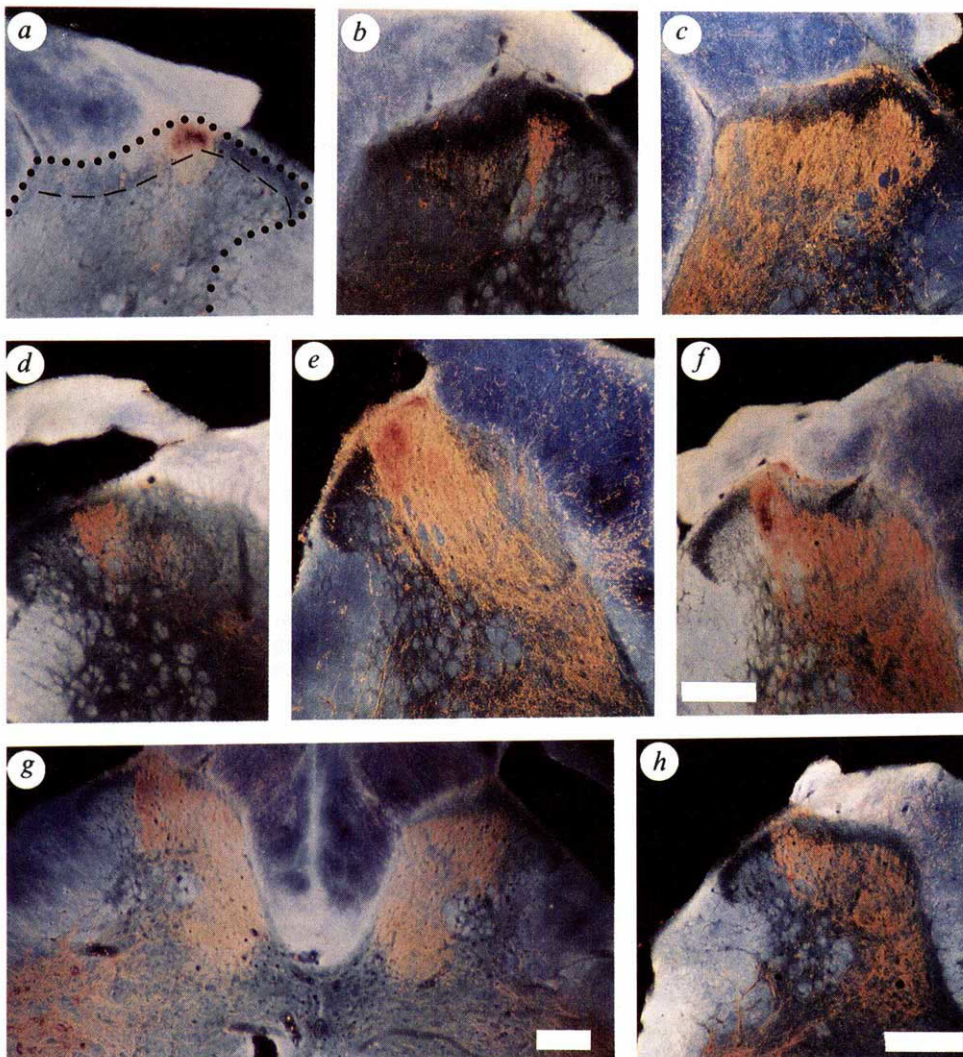


FIG. 1 Dark-field photomicrographs of transganglionic HRP label in the dorsal horn of the spinal cord seen in transverse sections. *a*, WGA-HRP applied to a control sural nerve and labelling predominantly C afferents in the sural region of lamina II in the fourth lumbar segment. The dots indicate the outline of the grey matter of the dorsal horn, the dashed lines show the lamina II/III border. *b*, CB-HRP applied to a control sural nerve, labelling myelinated axons in lamina III. *c*, CB-HRP label of a control sciatic nerve showing absence of label in the entire mediolateral extent of its distribution in lamina II. *d*, CB-HRP label of a sural nerve 1 week after its section and ligation. Note the dorsal extension of the label into lamina II. *e*, CB-HRP label of a sural nerve 2 weeks after its section and ligation, the entire lamina II is now labelled. *f*, CB-HRP label to the sciatic nerve of animals with a chronic sural nerve crush (4 weeks). At this stage the pattern is very similar to that found for a cut nerve with the dorsal extension in lamina II restricted to the sural terminal region. Scale bar (*a-f*), 200 μ m. *g*, CB-HRP applied to sciatic nerves ipsi- and contralateral to a chronic sural nerve section (9 weeks). *h*, CB-HRP label 15 weeks post-sural section. Scale bars, 200 μ m.

METHODS. Sural nerve section or crush performed in adult (200–250 g) rats anaesthetized with fentanyl/diazepam. After recovery bulk labelling was done under fentanyl/diazepam anaesthesia by intraneural injection of 1 μ l 2.5% WGA-HRP (Sigma) or 1.5% CB-HRP (List) through fine glass pipettes. 48–72 h later, the animals were terminally anaesthetized, perfused with fixative and the HRP reaction product demonstrated using the TMB technique as described²⁶.

be the consequence of high-threshold A δ afferents sprouting from lamina I, low-threshold down hair A δ afferents from lamina III and low-threshold A β afferents from lamina III/IV. To examine whether A β afferents participate in the reorganization of terminals, we filled 35 single chronically axotomized afferents (6–9 weeks) intra-axonally with HRP and compared their distribution with 13 intact sural hair follicle afferents recorded in control animals. The axotomized afferents had conduction velocities in the A β range, $17\text{--}61\text{ m s}^{-1}$ (31.7 ± 1.1 , s.e.m.), and

were driven by stimulation of the sural nerve at intensities less than $50\text{ }\mu\text{A}$ at $50\text{ }\mu\text{S}$. The absence of a peripheral receptive field prevented us from ascertaining to which functional class an individual axotomized afferent belonged, but three different groups of axons defined in terms of the general appearance of their terminal arbor were observed. These groups were those with arbors that resembled the flame-shaped appearance that is characteristic of hair follicle afferents^{2,3,30} ($n = 23$), those with dorsal and ventral projections resembling glabrous/hairy rapidly adapting afferents³ ($n = 4$), and those with deep ventrally projecting branches with an appearance similar to that of slowly adapting afferents³⁰ ($n = 8$). The high proportion of hair follicle-like afferents found is in keeping with density of these afferents in hairy skin⁶. The appearance of the terminal arbors from two control unaxotomized hair follicle afferent fibres is shown in Fig. 2a and b. None of the 119 terminal arbors from 13 control sural hair follicle afferent fibres penetrated outer lamina II (II_o) or more dorsally, whereas 10 (8.4%) terminated in inner lamina II (II_i). Of the 93 bouton containing terminal arbors from 23 chronically axotomized sural afferents with a flame-shaped appearance, however, 18 arbors from 12 different axons reached lamina II_i (19%), six arbors from four axons branched into lamina II_o (6.5%), and seven arbors from four axons terminated in lamina I (7.5%) (Fig. 2d). Another 32% of these flame-shaped arbors from axotomized fibres showed some feature not seen in controls; dorsally, ventrally and mediolaterally directed branches in an atypical pattern (Fig. 2c) and a loss of somatotopic register in the rostrocaudal axis. The single cell analysis was compatible with the bulk labelling results but may have underestimated the amount of sprouting because of a sampling bias towards large axons.

Our results show that peripheral nerve injury results in a redistribution of the central terminals of myelinated afferents. If the low-threshold mechanoreceptive afferent terminals that sprout into lamina II establish functional contacts with cells that normally would have a monosynaptic C-nociceptor input, this could lead to inappropriate responses to innocuous peripheral stimuli. Although many cells in the superficial laminae have an A β input, this is usually polysynaptic, and there is a population of high-threshold nociceptive-specific cells which do not have a low-mechanothreshold receptive field¹. Sprouting of afferents into novel territories might produce changes in the timing and strength as well as the distribution of synaptic inputs to dorsal horn neurons, with consequent modifications in sensory processing. Therefore, the already well demonstrated peripheral regenerative capacity of primary afferent neurons might be accompanied by an equally important central regenerative capacity, or plasticity, that in certain circumstances could be maladaptive. This may explain why neuropathic pain is often intractable and resistant to peripheral treatments. □

Received 30 July; accepted 9 October 1991.

1. Brown, A. G. *Organization in the Spinal Cord* (Springer, Heidelberg, 1981).
2. Brown, A. G., Rose, P. K. & Snow, P. J. *J. Physiol., Lond.* **272**, 779–797 (1977).
3. Shortland, P., Woolf, C. J. & Fitzgerald, M. *J. comp. Neurol.* **289**, 416–433 (1989).
4. Light, A. R. & Perl, E. R. *J. comp. Neurol.* **186**, 133–150 (1979).
5. Sugiyama, Y., Lee, C. L. & Perl, E. R. *Science* **234**, 358–361 (1986).
6. Lynn, B. & Carpenter, S. E. *Brain Res.* **238**, 29–43 (1982).
7. Ralston, H. J. & Ralston, D. D. *J. comp. Neurol.* **184**, 643–684 (1979).
8. Cambell, J. N., Raja, S. N., Meyer, R. A. & McKinnon, S. E. *Pain* **32**, 89–94 (1988).
9. Price, D. D., Bennett, G. J. & Raffii, M. *Pain* **36**, 273–288 (1989).
10. Scadding, J. W. in *Textbook of Pain* (eds Wall, P. D. & Melzack, R.) 522–534 (Churchill Livingstone, Edinburgh, 1989).
11. Devor, M. *Br. med. Bull.* **47**, 619–630 (1991).
12. Bennett, G. J. in *Dahlem Workshop on Towards a New Pharmacotherapy of Pain* (eds Basbaum, A. I. & Besson, J.-M.) 365–379 (Wiley, Chichester, 1991).
13. Skene, H. J. P. *A. Rev. Neurosci.* **12**, 127–156 (1989).
14. Skene, H. J. P. & Willard, M. *J. Cell Biol.* **89**, 96–103 (1981).
15. Hoffman, P. *J. Neurosci.* **9**, 893–897 (1989).
16. Sommerville, T., Reynolds, M. L. & Woolf, C. J. *Neuroscience* **45**, 213–220 (1991).
17. Tetzlaff, W., Zwiers, H., Lederis, K., Cassar, L. & Bisby, M. A. *J. Neurosci.* **9**, 1303–1313 (1989).
18. Woolf, C. J. *et al. Neuroscience* **34**, 465–478 (1990).
19. Coggeshall, R. E., Reynolds, M. L. & Woolf, C. J. *Neurosci. Lett.* **131**, 37–41 (1991).
20. Castro-Lopes, J. M., Coimbra, A., Grant, G. & Arvidsson, J. *J. Neurocytol.* **19**, 329–337 (1990).
21. Knyihar-Csillik, E., Rakic, P. & Csillik, B. *Cell Tissue Res.* **247**, 599–604 (1987).

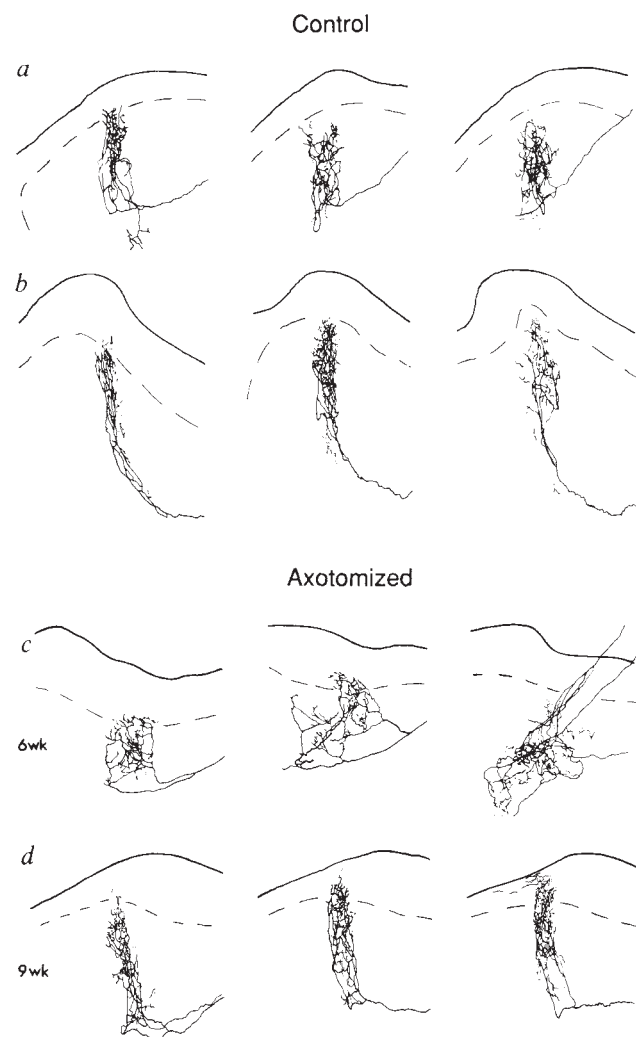


FIG. 2 Camera lucida reconstructions from transverse sections of the spinal cord of the complex terminal arbors of four single sural A β afferent fibres. Each row shows three adjacent arbors from a single afferent, two controls (a and b) and two from axotomized sural nerves (c and d). Solid line above each drawing, dorsal-most surface of the dorsal horn; dotted line, the lamina II/III border. Each arbor has a collateral axon arising from a stem axon running longitudinally in the dorsal columns (not shown). This collateral axon, for hair follicle afferent fibres, terminates in a 'flame-shaped' branching pattern (a and b), with mediolaterally restricted branches extending dorsally from lamina IV to the lamina II/III border and only a few sparse terminals in some cases reaching inner lamina II (refs 3, 30). c, Arbors which have a branching pattern not seen in control animals, a wide mediolateral spread, dorsal extension up to II_o and a disruption of the flame-shaped pattern. d, Arbors from an axotomized afferent which although retaining the flame-shape, extend in some cases right up to the marginal zone (lamina I). Scale bars, 250 μm .

METHODS. Intracellular recordings were made from control hair follicle afferents whose receptive fields fell in the sural nerve peripheral territory ($n = 13$) or from axotomized afferents which could be driven by electrical stimulation ($<50\text{ }\mu\text{A}$ at $50\text{ }\mu\text{S}$, 100 Hz , latency $<6\text{ ms}$) of axotomized sural nerves. The techniques for recording, dye injection, perfusion, histological processing and reconstruction have been described^{3,30}.

22. Sugimoto, T. & Gobel, S. *Brain Res.* **321**, 199–208 (1984).
23. McMahon, S. B. & Kett-White, R. *J. comp. Neurol.* **304**, 307–315 (1991).
24. Mesulam, M.-M. *Tracing Neural Connections with Horseradish Peroxidase* (Wiley, New York, 1982).
25. Brushart, T. M. & Mesulam, M.-M. *Neurosci. Lett.* **17**, 1–6 (1980).
26. Swett, J. E. & Woolf, C. J. *J. comp. Neurol.* **231**, 66–77 (1985).
27. Robertson, B. & Grant, G. *Neuroscience* **14**, 895–905 (1985).
28. Robertson, B., Perry, M. J. & Lawson, S. N. *J. Neurocytol.* **20**, 387–395 (1991).
29. Harper, A. A. & Lawson, S. N. *J. Physiol., Lond.* **359**, 31–46 (1985).
30. Woolf, C. J. *J. comp. Neurol.* **261**, 105–119 (1987).

ACKNOWLEDGEMENTS. This work was supported by the MRC, the Wellcome Trust, a Bristol-Myers Squibb Unrestricted Pain Research Grant, the ISRT and the NIH.

Engagement of the high-affinity IgE receptor activates *src* protein-related tyrosine kinases

Elisa Eiseman & Joseph B. Bolen

Department of Molecular Biology, Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey 08543-4000, USA

THE high-affinity IgE receptor (FcεRI), which is expressed on the surface of mast cells and basophils, has a central role in immediate allergic responses. In the rat basophilic leukaemia cell line RBL-2H3, which is a model system for the analysis of FcεRI-mediated signal transduction, surface engagement of FcεRI induces histamine release and the tyrosine phosphorylation of several distinct proteins¹. Although the α, β and γ subunits of FcεRI lack intrinsic tyrosine protein kinase (TPK) activity, a kinase that copurifies with FcεRI phosphorylates the β and γ subunits of the receptor on tyrosine residues^{2,3}. We report here that in RBL-2H3 cells, p56^{lyn} and pp60^{c-src} are activated after FcεRI crosslinking, and p56^{lyn} coimmunoprecipitates with FcεRI. In the mouse mast-cell line PT-18, another cell type used to study FcεRI-mediated signalling, tyrosine phosphorylation of proteins is also an immediate consequence of receptor crosslinking. Notably, the only detectable *src* protein-related TPK in PT-18 cells is p62^{c-yes}, and it is this TPK that is activated on FcεRI engagement and coimmunoprecipitates with the receptor. Therefore, it seems that different *src* protein-related TPKs can associate with the same receptor and become activated after receptor engagement.

RBL-2H3 cells were activated by two independent methods to evaluate signal transduction through FcεRI. RBL-2H3 cells were activated by antigen presentation (IgE + DNP) by first sensitizing the cells with anti-2,4-dinitrophenol (DNP) mouse monoclonal IgE and then stimulating with antigen. Alternatively, RBL-2H3 cells were activated by crosslinking FcεRI with antibodies directed against receptor surface components. Anti-phosphotyrosine (APT) immunoblotting revealed that stimulation of RBL-2H3 cells by both methods induced the phosphorylation of several distinct proteins on their tyrosine residues (Fig. 1a, b). Tyrosine phosphorylation was observed as early as 30 s after stimulation and continued to increase over 10 min.

The results shown in Fig. 1c demonstrate that a kinase was associated with components of the IgE receptor complex. FcεRI was immunoprecipitated from resting RBL-2H3 cells and the immunoprecipitates had *in vitro* kinase activity. Stimulation of RBL-2H3 cells with IgE + DNP or with anti-FcεRI antibodies, increased the kinase activity of the coimmunoprecipitating kinase. This apparent activation was observed as an increase in phosphorylation of a prominent doublet at relative molecular masses 53,000 and 56,000 (M_r , 53K and 56K), as well as proteins with M_r values consistent with those previously identified as the β (33K) and γ (12–14K) subunits of the receptor⁴.

As the *src* protein-related TPKs range in M_r from 53 to 62K, and several *src* protein-related TPKs associate with cell-surface receptors, such as p56^{lck} with CD4/CD8 (refs 5–7), p60^{lyn} with

the T-cell receptor⁸, p56^{blk}, p56^{lyn} and p60^{lyn} with the B-cell receptor⁹, and p60^{lyn}, p62^{c-yes} and p56^{lyn} with the platelet glycoprotein CD36 (ref. 10), the RBL-2H3 cells were assayed for the *in vitro* kinase activity of the *src* protein-related TPKs (Fig. 1d). The most abundant *src* protein-related TPK found in RBL-2H3 cells was p56^{lyn} and pp60^{c-src} activity was also detectable.

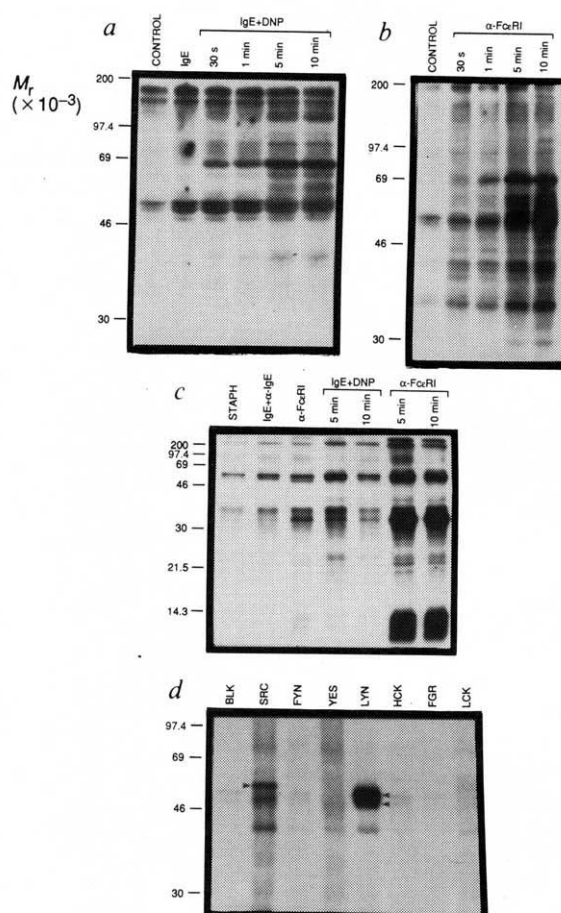
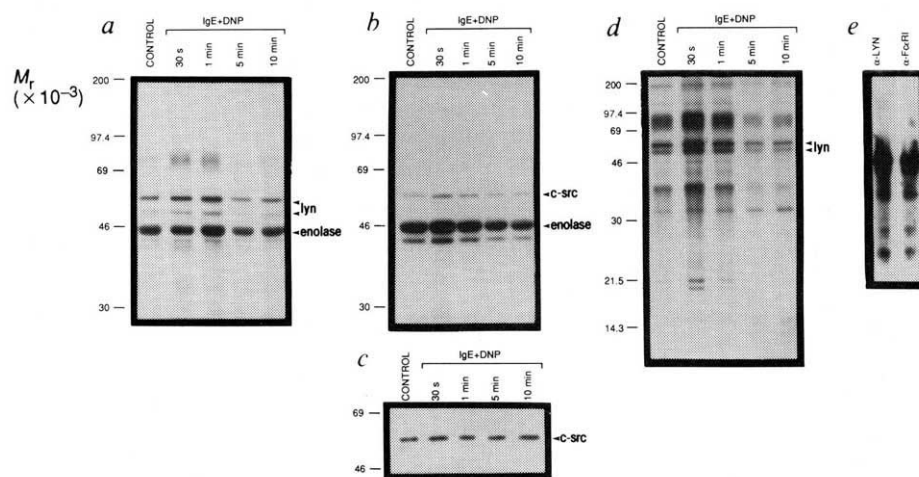


FIG. 1 Time course of tyrosine phosphorylation after activation and *src* protein-related TPKs in RBL-2H3 cells. *a*, APT immunoblot of RBL-2H3 lysates after activation with IgE + DNP (50 µg protein per lane). *b*, Anti-phosphotyrosine immunoblot of RBL-2H3 cells after activation with anti-FcεRI antibodies (50 µg protein per lane). *c*, FcεRI-associated immune-complex kinase activity after activation with IgE + DNP or anti-FcεRI antibodies. *d*, Expression of *src* protein-related TPKs in RBL-2H3 cells. The arrows indicate the positions of pp60^{c-src} and p53/p56^{lyn}. The positions of prestained M_r markers (Amersham) are indicated.

METHODS. RBL-2H3 cells have been previously described¹⁷. Cells were activated with IgE + DNP by first sensitizing the cells with 6 µg ml⁻¹ anti-DNP mouse monoclonal IgE from the hybridoma Hi-DNP-ε-26.82 (refs 18, 19) for 60 min at 37 °C, washed and then stimulated with 40 ng ml⁻¹ DNP-BSA at 37 °C for the indicated times before cell lysis. Resting cells (control) and cells sensitized with IgE alone (IgE) were also assayed. Cells were activated with the anti-FcεRI antibodies BC4 (ref. 20) (4 µg ml⁻¹) (*b*) and monoclonal antibody 5.14 (ref. 21) (6 µg ml⁻¹) (*c*) at 37 °C for the indicated times before cell lysis. Cells were solubilized at ~5 × 10⁷ cells ml⁻¹ in lysis buffer containing 10 mM CHAPS detergent, 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 10 µg ml⁻¹ aprotinin, 10 µg ml⁻¹ leupeptin, 100 µM sodium orthovanadate, and 5 mM NaF. Immunoblot analysis using rabbit anti-phosphotyrosine (Upstate Biotechnology), and immune-complex kinase assays have been previously described⁵. For immune-complex kinase assays, 250 µg total protein were used per reaction. FcεRI complexes from cells activated with IgE + DNP were immunoprecipitated with rabbit anti-mouse IgE (ref. 5). FcεRI complexes from cells activated with antibody 5.14 were immunoprecipitated with antibody 5.14. Monoclonal antibody 327 directed against pp60^{c-src} has been previously described²². Polyclonal anti-peptide antibodies to the other *src* protein-related TPKs were obtained from rabbits immunized with synthetic peptides corresponding to amino-acid sequences in the unique domain of each individual kinase²³.

FIG. 2 Time course of activation of p56^{lyn} and pp60^{c-src} in RBL-2H3 cells after stimulation with IgE + DNP. **a**, Immune-complex kinase assays after immunoprecipitation with anti-lyn protein antibodies. **b**, Immune-complex kinase assays after immunoprecipitation with anti-src protein antibodies. **c**, src protein immunoblot after src protein immunoprecipitation. **d**, Immune-complex kinase assays after immunoprecipitation with anti-FcεRI antibodies. The positions of p53/p56^{lyn}, pp60^{c-src}, and exogenous substrate rabbit muscle enolase are shown. **e**, Peptide mapping of total cellular p56^{lyn} and receptor-associated p56^{lyn}. **METHODS**. RBL-2H3 cells were activated with IgE + DNP as described in Fig. 1 for the indicated times, lysed and equal amounts of lysate (250 µg protein per reaction) used for the src protein immunoblot, and immune-complex kinase assays with or without rabbit muscle enolase as an exogenous substrate⁵. FcεRI was immunoprecipitated with a monoclonal antibody directed against the β subunit of the receptor (mAbβ-JRK)²⁴. Representative ³²P-labelled bands corresponding to those indicated by the arrows in Fig. 2a and d were



subjected to partial proteolytic peptide analysis using *Staphylococcus aureus* V8 protease.

In these cells, p56^{lyn} was expressed as a doublet with apparent M_r values of 53 and 56K. This doublet represents the products of the two alternatively spliced lyn transcripts previously described in murine myeloid cells¹¹.

To determine whether the activity of p56^{lyn} or pp60^{c-src} was altered after FcεRI engagement, the effects of activating through FcεRI with IgE + DNP on the kinase activity of total cellular p56^{lyn}, and total cellular pp60^{c-src}, were analysed (Fig. 2a, b). Both p56^{lyn} and pp60^{c-src} kinase activity increased in 30 s. This change in kinase activity seems to be an increase in specific activity as the abundance of pp60^{c-src} was not detectably altered in the course of the activation (Fig. 2c). We were unable to determine the abundance or specific activity of p56^{lyn} owing to the lack of appropriate immunologic reagents. But the rapid increase in p56^{lyn} activity after receptor engagement suggests that this change probably reflects an increase in specific activity.

Given the observed changes in p56^{lyn} and pp60^{c-src} activity after activation, coupled with the similarity in M_r of p56^{lyn} with the previously identified FcεRI-associated phosphoproteins (Fig. 1c), we evaluated whether p56^{lyn} or pp60^{c-src} could be coimmunoprecipitated with FcεRI. Immunoprecipitation of FcεRI revealed that the receptor-associated kinase was activated

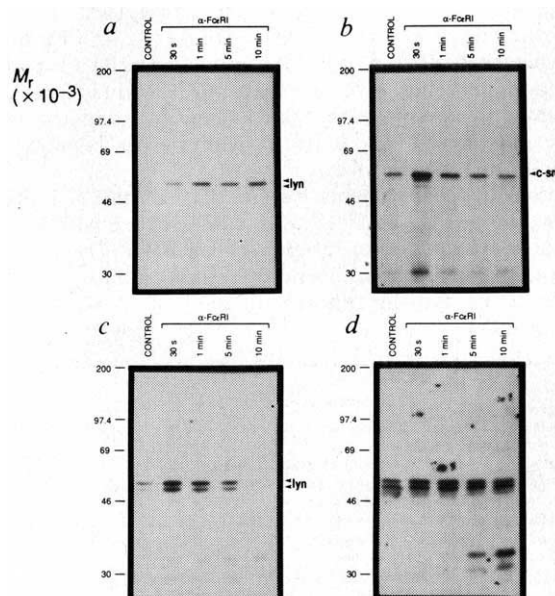
with the same kinetics as total cellular p56^{lyn} (Fig. 2d). Evaluation of the coimmunoprecipitating kinase by peptide mapping revealed that it was p56^{lyn} (Fig. 2e). In these experiments, there was no indication that pp60^{c-src} coimmunoprecipitated with FcεRI.

Direct crosslinking of FcεRI with anti-FcεRI antibodies also increased the *in vitro* kinase activity of total cellular p56^{lyn} and pp60^{c-src}, and receptor associated p56^{lyn} (Fig. 3). But the kinetics of activation of total cellular p56^{lyn} after direct crosslinking varied slightly from that observed after activation with IgE + DNP. Immunoprecipitation of the receptor with anti-FcεRI antibodies followed by APT immunoblotting revealed that the tyrosine phosphorylation state of p56^{lyn} paralleled the kinase activity of receptor-associated p56^{lyn} (Fig. 3d). We also observed an increase in APT reactivity of bands with M_r values similar to the β subunit of the receptor and a previously described 30K protein which may be another IgE-binding component of FcεRI (refs. 12–16).

The mouse mast-cell line, PT18, was used as an alternative system to study the role of the src protein-related TPKs in FcεRI-mediated signal transduction. As with RBL-2H3 cells, crosslinking of FcεRI with IgE + DNP induced the

FIG. 3 Time course of activation of p56^{lyn} and pp60^{c-src} in RBL-2H3 cells after stimulation with anti-FcεRI antibodies. **a**, Immune-complex kinase assays after immunoprecipitation with anti-lyn protein antibodies. **b**, Immune-complex kinase assays after immunoprecipitation with anti-src protein antibodies. **c**, Immune-complex kinase assays after immunoprecipitation with anti-FcεRI antibodies. **d**, APT immunoblot after immunoprecipitation with anti-FcεRI antibodies. The positions of p53/p56^{lyn} and pp60^{c-src} are shown.

METHODS. RBL-2H3 cells were activated with BC4 as described in Fig. 1 for the indicated times, lysed and equal amounts of lysate (250 µg protein per reaction) used for the APT immunoblot, and immune-complex kinase assays. FcεRI was immunoprecipitated with BC4.



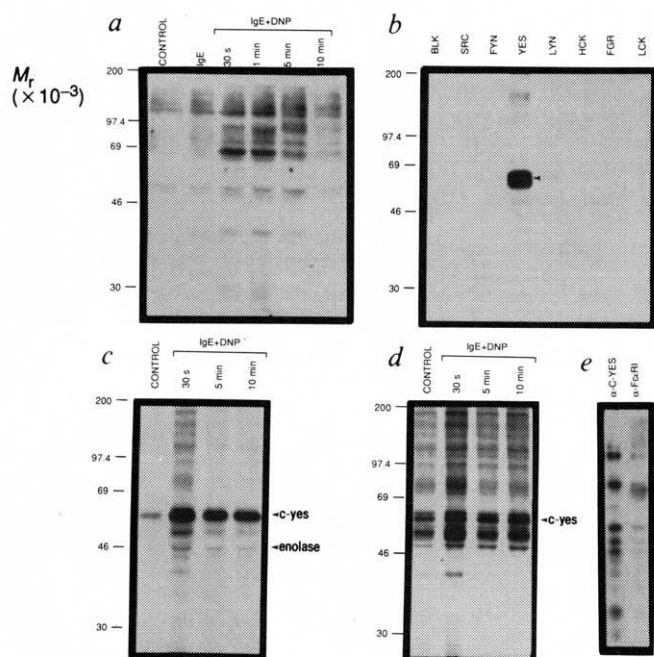


FIG. 4 Time course of activation of PT18 cells after stimulation with IgE + DNP. *a*, APT immunoblots of PT18 cell lysates after activation with IgE + DNP (50 µg protein per lane). *b*, Expression of *src* protein-related TPKs in PT18 cells. *c*, Immune-complex kinase assays after immunoprecipitation with anti-yes protein antibodies. *d*, Immune-complex kinase assays after immunoprecipitation with anti-FcεRI antibodies. The positions of p62^{c-yes} and exogenous substrate rabbit muscle enolase are shown. *e*, Peptide mapping of total cellular p62^{c-yes} and receptor associated p62^{c-yes}. METHODS. PT18 cells have been previously described²⁵. PT18 cells were activated with IgE + DNP as described in Fig. 1 for the indicated times, lysed and equal amounts of lysate (250 µg protein per reaction) used for immune-complex kinase assays with or without rabbit muscle enolase as an exogenous substrate⁵. FcεRI was immunoprecipitated with mAbβ-JRK (ref. 24). Representative ³²P-labelled bands corresponding to those indicated by the arrows in *c* and *d* were subjected to partial proteolytic peptide analysis using *S. aureus* V8 protease.

phosphorylation of several proteins on tyrosine, although with different kinetics (Fig. 4*a*). To determine whether the p56^{lyn} kinase was present in these cells, PT18 cells were assayed for the *in vitro* kinase activity of the *src* protein-related TPKs. Surprisingly, there was no detectable p56^{lyn} kinase activity. The only detectable *src* protein-related TPK in PT18 cells is p62^{c-yes} (Fig. 4*b*). Notably, p62^{c-yes} kinase activity was activated on crosslinking of FcεRI with IgE + DNP (Fig. 4*c*). Furthermore, immunoprecipitation of FcεRI with anti-FcεRI antibodies and peptide mapping revealed that p62^{c-yes} kinase activity co-immunoprecipitates with FcεRI (Fig. 4*e*), and the receptor-associated p62^{c-yes} is activated with similar kinetics to total cellular p62^{c-yes} (Fig. 4*d*).

From these experiments, it seems that both p56^{lyn}, in RBL-2H3 cells, and p62^{c-yes}, in PT18 cells, can associate with FcεRI and become activated after receptor engagement. These results are consistent with the concept that these kinases can participate in the tyrosine phosphorylation observed after FcεRI engagement. □

Received 5 July; accepted 20 September 1991.

- Benhamou, M., Gutkind, J. S., Robbins, K. C. & Siraganian, R. P. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5327-5330 (1990).
- Quarto, R. & Metzger, H. *Molec. Immun.* **23**, 1215-1223 (1986).
- Shimizu, A. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1907-1911 (1988).
- Metzger, H. et al. *A. Rev. Immun.* **4**, 419-470 (1986).
- Veilleux, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. *Cell* **55**, 301-308 (1988).
- Rudd, C. E., Trevillyan, J. M., Dasgupta, J. D., Wong, L. L. & Schlossman, S. F. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5190-5194 (1988).
- Barber, E. K., Dasgupta, J. D., Schlossman, S. F., Trevillyan, J. M. & Rudd, C. E. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3277-3281 (1989).

- Samelson, L. E., Phillips, A. F., Loung, E. T. & Klausner, R. D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4358-4362 (1990).
- Burkhardt, A. L., Brunswick, M., Bolen, J. B. & Mond, J. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7410-7414 (1991).
- Huang, M. M., Bolen, J. B., Barnwell, J. W., Shattil, S. J. & Brugge, J. S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7844-7848 (1991).
- Yi, T., Bolen, J. B. & Ihle, J. N. *Molec. cell. Biol.* **11**, 2391-2398 (1991).
- Hempstead, B. L., Parker, C. W. & Kulczycki, A. Jr. *J. biol. Chem.* **256**, 10717-10723 (1981).
- Holowka, D. & Baird, B. *J. biol. Chem.* **259**, 3720-3728 (1984).
- Kulczycki, A. Jr & Parker, C. W. *J. biol. Chem.* **254**, 3187-3193 (1979).
- Hempstead, B. L., Kulczycki, A. Jr & Parker, C. W. *Biochem. biophys. Res. Commun.* **98**, 815-822 (1981).
- Holowka, D., Hartmann, H., Kanelloupolous, J. & Metzger, H. *J. Receptor Res.* **1**, 41-68 (1980).
- Barsumian, E. L., Isersky, C., Petrino, M. G. & Siraganian, R. P. *Eur. J. Immun.* **11**, 317-323 (1981).
- Lui, F.-T. et al. *J. Immun.* **124**, 2728-2736 (1980).
- Holowka, D. & Metzger, H. *Molec. Immun.* **19**, 219-227 (1982).
- Basciano, L. K., Berenstein, E. H., Kmak, L. & Siraganian, R. P. *J. biol. Chem.* **261**, 11823-11831 (1986).
- Baniyash, M., Alkalay, I. & Eshhar, Z. *J. Immun.* **138**, 2999-3004 (1987).
- Lipsich, L. A., Lewis, A. J. & Brugge, J. S. *J. Virol.* **48**, 352-360 (1983).
- Bolen, J. B., Thompson, P. A., Eisenman, E. & Horak, I. D. *Adv. Cancer Res.* **57**, 103-149 (1991).
- Rivera, J., Kinet, J.-P., Kim, J., Pucillo, C. & Metzger, H. *Molec. Immun.* **25**, 647-661 (1988).
- Pluznik, D. H., Tare, N. S., Zatz, M. M. & Goldstein, A. L. *Exp. Hemat.* **10**, 211-218 (1982).

ACKNOWLEDGEMENTS. We thank H. Metzger and G. Alber for reagents and advice and R. Siraganian for the BC4 antibody.

Developmental switch of CREM function during spermatogenesis: from antagonist to activator

Nicholas S. Foulkes, Britt Mellström, Enrico Benusiglio & Paolo Sassone-Corsi*

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cedex, France

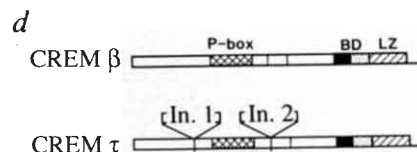
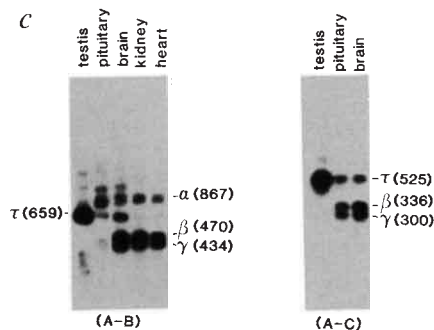
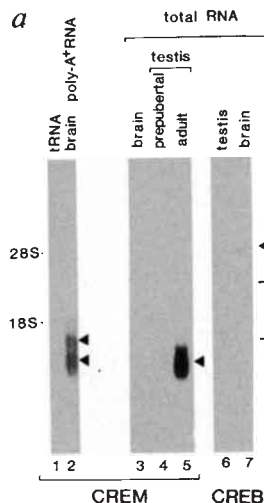
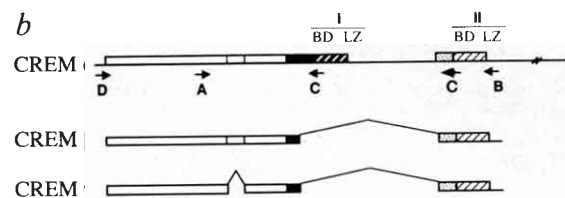
MAMMALIAN spermatogenesis consists of a series of complex developmental processes controlled by the pituitary-hypothalamic axis¹. This flow of biochemical information is directly regulated by the adenylate cyclase signal transduction pathway². We have previously described the CREM (cyclic AMP-responsive element modulator) gene which generates, by cell-specific splicing, alternative antagonists of the cAMP transcriptional response³. Here we report the expression of a novel CREM isoform (CREM τ) in adult testis. CREM τ differs from the previously characterized CREM antagonists by the coordinate insertion of two glutamine-rich domains that confer transcriptional activation function. During spermatogenesis there was an abrupt switch in CREM expression. In premeiotic germ cells CREM is expressed at low amounts in the antagonist form. Subsequently, from the pachytene spermatocyte stage onwards, a splicing event generates exclusively the CREM τ activator, which accumulates in extremely high amounts. This splicing-dependent reversal in CREM function represents an important example of developmental modulation in gene expression.

Figure 1*a* shows a northern blot of total RNA from adult testis, prepubertal (12-day-old) testis and brain. In the adult testis, a single intense CREM band of roughly 2 kilobases (kb) was visible; no bands were visible in the prepubertal testis or brain. This CREM transcript abundance is much higher than in brain where 5 µg poly(A)⁺RNA are required to detect CREM transcript (Fig. 1*a*, lane 2; and ref. 3). In contrast to CREM, a faint band of 7 kb corresponding to the transcript of nuclear factor CREB (an activator of cAMP-responsive promoter elements (CREs); refs 4, 5) was found in both brain and testis (lanes 6-7). Polymerase chain reaction (PCR) analysis using primers A and B (Fig. 1*b*) detected the three antagonistic isoforms: CREM α (867 base pairs (bp)), CREM β (470 bp) and

* To whom correspondence should be addressed.

FIG. 1 *a*, Northern blot analysis of CREM (lanes 1–5) and CREB (lanes 6 and 7) expression in poly-(A)⁺ (lane 2) and total RNA samples (lanes 3 to 7). Filters containing 5 µg of brain poly-(A)⁺ RNA (lane 2) or 10 µg total RNA (lanes 3–7) were hybridized with an *Nco*I–*Sma*I CREM α probe³ or a full-length CREB complementary DNA fragment⁵. Arrowheads, CREM-specific bands in brain, the single 2-kb CREM band visible in testis RNA, and the position of the 7-kb CREB transcript. *b*, Schematic representation of the previously described CREM isoforms: CREM α , β and γ (ref. 3). Shaded boxes denote the basic domain (BD) and leucine zipper (LZ) regions of the two alternative DNA binding domains (I) and (II). The four PCR primers (A, B, C and D) are indicated underneath. *c*, Southern blot analysis of RNA PCR reactions using primers A, B and C. The isoforms corresponding to each band and the size (bp) are indicated. Only single bands corresponding to the novel isoform τ are visible for testis RNA in both analyses. *d*, Schematic representation of the CREM β and τ isoforms. The cross-hatched box is the phosphorylation domain¹⁰ (P-box). The sites of two insertions in the τ isoform (In.1 and In.2) are indicated above.

METHODS. Total RNA was extracted from adult tissue as described¹⁷ and poly-(A)⁺ RNA was purified using a messenger RNA purification kit (Pharmacia). RNAs were northern blotted from 1% formaldehyde gels. Probes were labelled with [α -³²P]-dCTP using a Multiprime labelling kit (Amersham). Filters were hybridized in a buffer containing 50% formamide and the final wash was in 0.2×SSC, 0.1% SDS at 42 °C. PCR amplification of RNA was done as previously described³. The sequences of PCR primers used were: A: 5'-



GATTGAAGAAGAAAAATCAGA-3' (255–275), B: 5'-CATGCTGTAATCAGTTCATAG-3' (1121–1101), C: 5'-TTGACATATTCTTCTTCTT-3' (590–571) and D: 5'-GAGGACAAATGTAAGGCAAATGAC-3' (–19–5). Nucleotide positions correspond to the published CREM α sequence³.

CREM γ (434 bp) (Fig. 1*b*, *c*; and ref. 3). In testis a novel 659-bp fragment was amplified exclusively (Fig. 1*c*). This band was also detected in some neuroendocrine tissues, but at much lower abundance and in combination with the other CREM isoforms (pituitary, brain; Fig. 1*c*). Sequencing of the 659-bp fragment revealed that it corresponds to the CREM β form with a 189-bp insertion (In.2 in Fig. 1*d*; between positions 352 and 353 in the CREM α sequence; ref. 3). Consistent with this result we also observed a single fragment of 525 bp using primers A and C (Fig. 1*b*). Amplification of the complete coding region (primers B and D, Fig. 1*b*) generated a single fragment which included an additional insertion of 147 bp (In.1 in Fig. 1*d*). The resulting 1,023-bp open reading frame encodes a protein of 341 amino acids (Fig. 2*a*) very similar to CREB (Fig. 2*b*). The structure and abundance of this transcript have been confirmed by RNase protection (see legend to Fig. 2; and data not shown). Additional analysis demonstrated a limited distribution of the τ isoform in other tissues and cell lines, notably those of neuroendocrine origin; however, it is always found in combination with the other isoforms and in amounts comparable to brain (Fig. 1, and ref. 3; data not shown).

DNA binding of CREM τ protein was tested and specificity for recognition of a CRE was confirmed as predicted from previous studies^{3,6} (Fig. 3*a*, *b*). Next we tested whether CREM τ can regulate transcription. In transfection assays we showed that CREM τ , in contrast to the other CREM isoforms, functions as a transcriptional activator when coexpressed with protein kinase A (Fig. 3*c*, *d*; see also ref. 7). Consistent with our previous findings^{3,6}, CREM α and CREM β act as antagonists (Fig. 3*d*).

The two inserted amino-acid regions in CREM τ are at least three times richer in glutamine residues than the rest of the protein (16.3% glutamine in In.1 and 15.9% in In.2 compared with 5.6% in the remainder of the protein). Analogous domains in CREB share this feature and are involved in transactivation function⁸. Glutamine richness is also a feature of the activation domains of other transcription factors, for example Sp1 (ref. 9). Several models could explain how the inserted domains modify the structure and thereby the protein function, considering that phosphorylation at the P-box (Fig. 1*d*) is also thought to have been important¹⁰. The inserted domains may change the tertiary structure of the protein by increasing the exposure of the P-box, or they may be involved in direct contacts with other components of the transcriptional machinery. More important, the isoforms produced by the CREM gene constitute a model for alternative functionality generated by the natural shuffling of various protein domains.

We next examined at which stage in testis development the CREM switch occurs. We first analysed testis RNA from prepubertal mice of different ages¹¹. In 1-week-old mice, weak bands representing all CREM forms are obtained (Fig. 4*a*). But from 2 weeks onwards the pattern changes greatly, showing only the CREM τ form. The change in expression pattern occurs between 13 and 14 days (not shown). This switch coincides with the dramatic increase in the amount of CREM expression (Fig. 4*a*, and data not shown) and with the pachytene spermatocyte stage. RNA from 20-day-old irradiated mice shows the pre-switch pattern of CREM expression confirming the germ-cell-based expression of the abundant CREM τ form¹² (Fig. 4*a*, lanes

a

Met Thr Met Glu Thr Val Glu Ser Gln Gln Asp Arg Ser Val Thr Arg Ser Val Ala Glu 20
 1 ATG ACC ATG GAA ACA GTT GAA TCA CAG CAG GAT CGA AGT GTA ACA CGT TCT GTG GCA GAG

His Ser Ser Ala His Met Gln Thr Gly Gln Ile Ser Val Pro Thr Leu Ala Gln Val Ser 40
 61 CAT AGC TCT GCT CAT ATG CAG ACT GGT CAA ATT TCT GTT CCT ACT CTA GCT CAG GTT TCT

Val Ala Gly Ser Gly Thr Gly Arg Gly Ser Pro Ala Val Thr Leu Val Gln Leu Pro Ser 60
 121 GTA GCT GGA TCA GGC ACT GGA AGA GGC TCC CCA GCT GTG ACT CTA GTA CAG TTA CCT TCA

Gly Arg Thr Val Gln Val Gln Gly Val Ile Gln Thr Pro His Pro Ser Val Ile Gln Ser 80
 181 GGC CAA ACT GTA CAG GTC CAG GGA GTT ATT CAG ACA CCA CAT CCA TCG GTT ATT CAA TCA

Pro Gln Ile Gln Thr Val Gln Val Ala Thr Ile Ala Glu Thr Asp Asp Ser Ala Asp Ser 100
 241 CCA CAA ATA CAA ACT GTT CAG GTA GCA ACA ATT GCA GAG ACA GAT GAT TCT CCA GAC TCA

Glu Val Ile Asp Ser His Lys Arg Arg Glu Ile Leu Ser Arg Arg Pro Ser Tyr Arg Lys 120
 301 GAA GTA ATT GAT TCG CAT AAA CGT AGA GAA ATT CTT TCA CGA AGA CCC TCA TAT AGA AAA

Ile Leu Asn Glu Leu Ser Ser Asp Val Pro Gly Ile Pro Lys Ile Glu Glu Glu Lys Ser 140
 361 ATA CTG AAT GAA CTT TCC TCT GAT GTG CCT GGT ATT CCC AAG ATT GAA GAA GAA AAA TCA

Glu Glu Glu Gly Thr Pro Pro Asn Ile Ala Thr Met Ala Val Pro Thr Ser Ile Tyr Gln 160
 421 GAG GAA GAA GGG ACA CCA CCT AAC ATT GCT ACC ATG GCA GTA CCA ACT AGC ATA TAT CAG

Thr Ser Thr Gly Gln Tyr Ile Ala Ile Ala Gln Gly Gly Thr Ile Gln Ile Ser Asn Pro 180
 481 ACT AGC ACG GGG CAA TAC ATT GCT ATA GCT CAA GGT GGA ACA ATC CAG ATT TCT AAC CCA

Gly Ser Asp Gly Val Gln Gly Leu Gln Ala Leu Thr Met Thr Asn Ser Gly Ala Pro Pro 200
 541 GGA TCT GAT GGT GTT CAG GGA CTC CAG GCA TTA ACA ATG ACA ATC TCA GGA GCT CCT CCG

Pro Gly Ala Thr Ile Val Gln Tyr Ala Ala Gln Ser Ala Asp Gly Thr Gln Gln Phe 220
 601 CCA GGT GCT ACA ATT GTA CAG TAT GCA GCA CAA TCA GCC GAT GGT ACA CAG CAG TTC TTT

Val Pro Gly Ser Gln Val Val Val Gln Asp Glu Glu Thr Asp Leu Ala Pro Ser His Met 240
 661 GTC CCA GGC AGC CAG GTT GTT GTT CAA GAT GAG GAG ACT GAC CTT GCC CCA AGT CAC ATG

Ala Ala Ala Thr Gly Asp Met Pro Thr Tyr Gln Ile Arg Ala Pro Thr Thr Ala Leu Pro 260
 721 GCT GCT GCC ACA GGT GAC ATG CCA ACT TAC CAG ATC CGA GCT CCT ACT ACT GCT TTG CCA

Gln Gly Val Val Met Ala Ala Ser Pro Gly Ser Leu His Ser Pro Gln Gln Leu Ala Glu 280
 781 CAA GGT GTG GTG ATG GCT GCC TCA CCA GGA AGC CAG AGT CCC CAG CTA GCA GAA

Glu Ala Thr Arg Lys Arg Glu Leu Arg Leu Met Lys Asn Arg Glu Ala Ala Lys Glu Cys 300
 841 GAA GCA ACT CGC AAG CCG GAG CTG AGG CTG ATG AAA AAC AGG GAA GCT GCT AAA GAA TGT

Arg Arg Arg Lys Lys Glu Tyr Val Lys Cys Leu Glu Ser Arg Val Ala Val Leu Glu Val 320
 901 CGA CGT CGA AAG AAA GAG TAT GTG AAG TGT CTT GAG AGT CGA GTC GCA GTG CTG GAA GTT

Gln Asn Lys Lys Leu Ile Glu Glu Leu Glu Thr Leu Lys Asp Ile Cys Ser Pro Lys Thr 340
 961 CAG AAC AAG AAG CTT ATA GAG GAG CTT GAA ACT TTG AAA GAC ATT TGC TCT CCC AAA ACA

Asp *** 341
 1021 GAT TAG

b

CREB	MTMESGAENQQSGDAAVTEAENQOMTVQAQP-QI--ATLAQVSNPAAHAT	47
CREM τ	MTMETV-ESQQ--DRSVTRSVAEHSSAHMQTGQISVPTLAQVSVAGSGTG	47
CREM β	MTMETV-ESQQ--DRSVTRSVAEHSSAHMQTGQISVPTLAQ-----	38
CREB	SSAPVTTLVRCFPGNNS-QVHGVIQAAQPSVLIQSPQVQTVOISTIAESDE	96
CREM τ	RGSPAVTLVQLPSGRTVQVQGVGTTPHEVLIQSPQVGTVOVATIAETD	97
CREM β	-----VATIAETD	48
CREB	QESVDSVTDSQKRREILSRPSYRKILNDESSDAPGVPIREEKSEET	146
CREM τ	ADSE--VIDSHKRREILSRPSYRKILNELSSDVPGIPKIEEKSEEGT	145
CREM β	ADSE--VIDSHKRREILSRPSYRKILNELSSDVPGIPKIEEKSEEGT	96
CREB	APAITVTVPTPIYQTSSQYIAITQGGAIQLANNGTDGVQGLQTLMTN	196
CREM τ	PPNIATMAVPTSIYQTSTGQYIAIAQGGTIQISNPGSDGVQGLQALMTN	195
CREM β	PPNIATMAVPTSIYQTSTGQY-----	117
CREB	AAATPGTTLQYA-QTTDG-QQILVPSNQVVQ-----AASG	232
CREM τ	SGAPPEGATIVQYAAQSADGTQQFFVPGSQVVQDEETDLAPSHMAAATG	245
CREM β	-----NEETDLAPSHMAAATG	133
CREB	DSQTYQIRTAAPTSTIAPGVVMASSPAL---PTQPAEEAARKREVRLMKNR	279
CREM τ	DMPTYQIR-APTALPQGVVMASSPGSLHSPQQLAEEATRKRRLMKNR	294
CREM β	DMPTYQIR-APTALPQGVVMASSPGSLHSPQQLAEEATRKRRLMKNR	184
CREB	EAAKECRRRKRYVKCLESRVAVLEVNKKLIEELETLDKIDCSPQTD	326
CREM τ	EAAKECRRRKRYVKCLESRVAVLEVNKKLIEELETLDKIDCSPQTD	341
CREM β	EAAKECRRRKRYVKCLESRVAVLEVNKKLIEELETLDKIDCSPQTD	229

6-8). We tested expression in testicular feminization (Tfm) and oligotriche (*olt*) mutants^{13,14}. PCR and northern analyses demonstrated normal adult amounts of CREM τ in *olt/olt* mice but prepubertal levels in the Tfm/Y samples (Fig. 4b; lanes 1-3). The amount of CREB in both mutants is equivalent to that in normal adult testis (Fig. 4b, lanes 4-6). Finally, in

fractionated testis cells the CREM τ form is expressed strongly in spermatocytes and spermatids, whereas the pre-switch pattern is present in both Sertoli and Leydig cells (Fig. 4c).

The striking developmental switch from transcriptional antagonist to activator represents an important example of a developmentally regulated splicing event which generates alternative

FIG. 2 a, The 1,023-bp cDNA sequence encoding the CREM τ isoform. The predicted 341 amino-acid open reading frame encodes a protein with a relative molecular mass (M_r) of 36,000. Nucleotide 1 corresponds to the first A of the initiation methionine. Dashed boxes containing amino-acids 105-109, 114-120 and 136-143 indicate sequences conserved in CREB that are putative phosphorylation sites for protein kinase C, protein kinases A and C and casein kinase II, respectively^{4,5,10}. The DNA-binding domain and the individual leucines of the leucine zipper are boxed. Insertions 1 and 2 (147 bp and 189 bp) are underlined (dashed) and delineated by arrows. The splicing pattern of the CREM τ transcript was confirmed by RNase protection analysis using three probes³: one straddles the 5' border of DNA-binding domain II in CREM α (ref. 3); in addition a probe which extends from a *PvuII* site in insertion 1 (position 154) to the 5'-end of the cDNA clone and a third probe extending from a *SacI* site in insertion 2 (position 590) to an *NcoI* site (position 452). b, Amino-acid sequence similarity between CREM τ and CREB. The predicted 341 amino-acid sequence of CREM τ is aligned with the 326 amino-acid sequence of human placental CREB (ref. 4) together with the 229 amino-acid sequence of CREM β (ref. 3). Shaded regions show sequence identity between CREM τ and CREB. Gaps introduced in the sequences to optimize the alignment are represented by dashes. Only CREM τ amino-acids 230-241 have no counterpart in the CREB sequence and correspond to the sequence which is deleted from the CREM γ isoform³.

METHODS. Alkali-denatured plasmid DNA was sequenced using T7 DNA polymerase (Sequenase kit; USB) in conjunction with custom-synthesized 20- and 21-base oligonucleotide primers (complementary to the cDNA sequence).

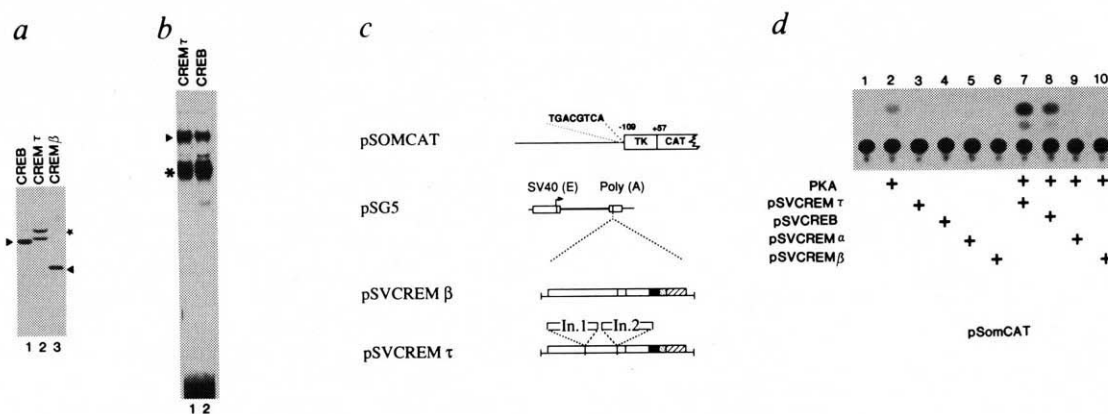
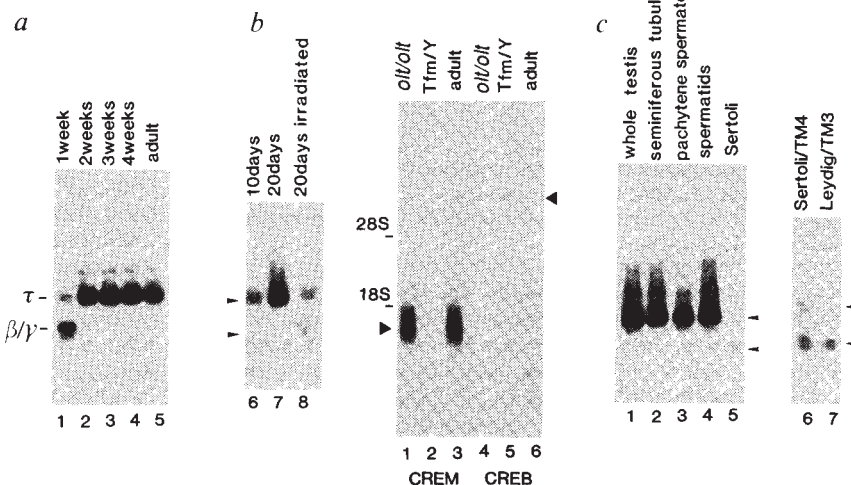


FIG. 3 CREM τ is a CRE-specific transcriptional activator. *a*, CREB, CREM τ and CREM β proteins are generated *in vitro* in a reticulocyte lysate system in the presence of 35 S-methionine and visualized on a 10% polyacrylamide-SDS gel. The apparent M_r of the proteins corresponds to the prediction from the deduced amino-acid sequence (see Fig. 2). The star indicates a nonspecific translation product of a lysate endogenous mRNA³. *b*, Representative gel retardation analysis of *in vitro* synthesized CREB and CREM τ proteins on a canonical CRE site. Arrowhead, specific protein complex; asterisk, nonspecific complex formed by lysate endogenous proteins¹⁸. Additional experiments already demonstrated that CREB and CREM protein binding is CRE-specific^{3,6}. *c*, The reporter plasmid pSomCAT has been used to score for the regulation exerted on a canonical CRE sequence TGACGTCA, inserted upstream from the herpes thymidine kinase promoter (tk)³. Complete CREM α , β and τ cDNAs and the rat CREB cDNA⁵ were cloned in the pSG5 expression vector³. *d*, Representative result of a JEG-3 transfection

experiment. Equivalent data were obtained using HeLa, F9 and NIH3T3 cells (not shown). Activation of pSomCAT transcription is obtained by cotransfection of the protein kinase A (PKA) catalytic subunit expression vector pC α EV (compare lanes 1 and 2; ref. 7). Transfection of CREM α , CREM β , CREM τ or CREB expression vectors in the absence of PKA does not affect pSomCAT expression (lanes 3–6). As already reported, CREM α and CREM β antagonize CRE-dependent activation^{3,6} (lanes 9–10). CREB and CREM τ augment the activation obtained by PKA (lanes 7 and 8).

METHODS. Linearized plasmid from CREB and CREM cDNA subclones in pGEM4 (Promega) were transcribed *in vitro* using T7 RNA polymerase as described previously¹⁸. Purified RNA was translated *in vitro* in a micrococcal nuclease-treated, methionine-free rabbit reticulocyte lysate (Promega). Gel retardation assays were done as described¹⁸. Cells were cultured and transfected as described previously³.

FIG. 4 *a*, PCR analysis of prepubertal testis CREM mRNA using primers A and C (Fig. 1). Lanes 1–4, total RNA from the testis of 1-, 2-, 3- and 4-week-old mice. Lane 5, adult testis positive control. Between weeks 1 and 2, the expression of CREM switches abruptly from encoding predominantly the transcriptional antagonists (β/γ in lane 1) to almost exclusively the activator (τ in lanes 2–5). Lanes 6–8, equivalent analysis of CREM mRNA in irradiated testis. Lanes 6 and 7, normal animals of 10- and 20-days-old, respectively. Lane 8, 20-day-old animal irradiated *in utero*. *b*, Northern blot analysis of mutant mice testis RNA. Aliquots (10 μ g) of total testis RNA were analysed. Samples from adult mice mutants; oligotriche (*olt/olt*) (lanes 1 and 4) and testicular feminization (*Tfm/Y*) (lanes 2 and 5) and from a normal adult control (adult) (lanes 3 and 6) were included. Duplicate filters were hybridized with CREM- and CREB-specific probes (lanes 1–3 and lanes 4–6, respectively). CREM- and CREB-specific bands are indicated (see also Fig. 1). *c*, CREM expression in RNAs from isolated testis cells. Total RNA was prepared from pachytene spermatocytes, spermatids and Sertoli cells. In addition RNA was prepared from isolated seminiferous tubules and the mouse Sertoli and Leydig cell lines, TM3 and TM4 (ref. 19). Aliquots (0.5 μ g) were tested by PCR using primers A and C (see *a* and Fig. 1). RNA from adult testis was included as a control (whole testis). METHODS. Oligotriche mice (*olt/olt*) were obtained from the Centre de



Service des Animaux de Laboratoire (CNRS, Orléans). We used the C57/Black6 strain as wild-type mice in these studies. Mice were irradiated *in utero* with 100 rads for 20 s at 19 days postcoitum¹². Spermatogenic cells were prepared from adult mouse testes by unit gravity sedimentation on linear gradients of bovine serum albumin as described²⁰.

forms of a transcription factor with opposite functions. But it also raises the question as to the role of this abundant transactivator in testis, particularly when CREB continues to be expressed at normal, low amounts. The expression of testis-specific isoforms of the catalytic and regulatory subunits of protein kinase A (refs 15, 16) implies that in spermatogenic cells the cAMP signal transduction pathway needs to fulfil a special-

ized function and CREM τ may represent a nuclear target of this pathway. An alternative hypothesis could be that CREM τ is actually incorporated into the structure of the sperm head and has a role after fertilization. The discovery of CREM τ provides a new dimension to the transcriptional regulation exerted by CREM. Its expression pattern is also a clue to an important physiological role for this factor. □

Received 17 July; accepted 24 October 1991.

1. Santen, R. J. in *Endocrinology and Metabolism* (eds Felig, P., Baxter, J. D., Broadus, A. E. & Frohman, L. A.) 821-905 (McGraw-Hill, New York, 1987).
2. Ewing, L. L. & Robaire, B. *Ann. N.Y. Acad. Sci.* **564**, 1-302 (1989).
3. Foulkes, N. S., Borrelli, E. & Sassone-Corsi, P. *Cell* **64**, 739-749 (1991).
4. Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L. & Habener, J. F. *Science* **242**, 1430-1433 (1988).
5. Gonzalez, G. A. *et al.* *Nature* **337**, 749-752 (1989).
6. Foulkes, N. S., Laoidé, B. M., Schlotter, F. & Sassone-Corsi, P. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5448-5452 (1991).
7. Mellon, P. L., Clegg, C. H., Correll, L. A. & McKnight, G. S. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4887-4891 (1989).
8. Gonzales, G. A., Menzel, P., Leonard, J., Fischer, W. & Montminy, M. R. *Molec. cell. Biol.* **11**, 1306-1312 (1991).
9. Couray, A. J. & Tjian, R. *Cell* **55**, 887-898 (1989).
10. Lee, C. Q., Yun, Y., Hoeffler, J. P. & Habener, J. F. *EMBO J.* **9**, 4455-4465 (1990).
11. Willison, K. & Ashworth, A. *Trends Genet.* **3**, 351-355 (1987).
12. Dym, M. in *Histology Cell and Tissue Biology* (ed. Weiss, L.) 1000-1053 (Elsevier Biomedical, New York, 1983).
13. Lyon, M. F. & Hawkes, S. G. *Nature* **227**, 1217-1219 (1970).
14. Moutier, R. in *The Laboratory Animal in the Study of Reproduction* (eds Antikatzides, T., Erichsen, S. & Spiegel, A.) 5-7 (Fischer, Stuttgart, 1976).
15. Beebe, S. J. *et al.* *Molec. Endocrin.* **4**, 465-475 (1990).
16. McKnight, G. S. *et al.* *Recent Prog. Horm. Res.* **44**, 307-335 (1988).
17. Auffray, C. & Rougeon, F. *Eur. J. Biochem.* **107**, 303-314 (1980).
18. Sassone-Corsi, P., Ransone, L. J., Lamph, W. W. & Verma, I. M. *Nature* **336**, 692-694 (1988).
19. Mather, J. P. *Biol. Reprod.* **23**, 243-251 (1980).
20. Kilpatrick, D. L., Borland, K. & Jin, D. F. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5695-5699 (1987).

ACKNOWLEDGEMENTS. We thank E. Borrelli, J. R. Naranjo, B. Laoidé, D. Masquillier and J. P. Montmayeur for discussions, C. Passananti, M. Russo and P. Amati (University of Rome) for RNA samples from fractionated testis cells, Ph. Avner (Institut Pasteur) for the Tfm/Y mice, S. G. McKnight (Washington University) for pCαEV, J. Auwerx, P. Pevet, B. Laoidé, D. Masquillier, M. Acker and J. P. Montmayeur for materials, F. Schlotter for technical assistance and B. Boulay and F. Haenel for help in the preparation of the manuscript. B.M. was on leave from the Cajal Institute for Neuroscience, Madrid, N.S.F. was supported by a long-term EMBO fellowship. This work was supported by the CNRS, INSERM, Association pour la Recherche contre le Cancer and Rhone-Poulenc-Rorer.

Mapping of the protein import machinery in the mitochondrial outer membrane by crosslinking of translocation intermediates

Thomas Söllner, Joachim Rassow, Martin Wiedmann*†, Jens Schlossmann, Petra Keil, Walter Neupert & Nikolaus Pfanner‡

Institut für Physiologische Chemie, Universität München, Goethestrasse 33, W-8000 München 2, Germany

* Zentralinstitut für Molekularbiologie der Akademie der Wissenschaften, O-1115 Berlin-Buch, Germany

MITOCHONDRIA contain a complex machinery for the import of nuclear-encoded proteins^{1,2}. Receptor proteins exposed on the outer membrane surface are required for the specific binding of precursor proteins to mitochondria, either by binding of cytosolic signal recognition factors or by direct recognition of the precursor polypeptides¹⁻⁵. Subsequently, the precursors are inserted into the outer membrane at the general insertion site GIP (general insertion protein)⁶⁻¹⁰. Here we report the analysis of receptors and GIP by crosslinking of translocation intermediates and by coimmunoprecipitation. Surface-accumulated precursors were cross-linked to the receptors MOM19 and MOM72, suggesting a direct interaction of preproteins with surface receptors. We identified three novel mitochondrial outer membrane proteins, MOM7, MOM8, and MOM30 that, together with the previously identified MOM38, seem to form the GIP site and are present in the mitochondrial receptor complex.

The precursor of the ADP/ATP carrier, the most abundant protein of the mitochondrial inner membrane, was synthesized in rabbit reticulocyte lysate in the presence of ³⁵S-methionine¹¹. To accumulate the precursor at the mitochondrial outer mem-

brane (receptor stage and GIP stage), isolated *Neurospora crassa* mitochondria were incubated with the precursor in the presence of ATP, but in the absence of a membrane potential across the inner membrane¹¹⁻¹³. The mitochondria were incubated with the homobifunctional crosslinking reagent DSS (disuccinimidyl suberate) or the heterobifunctional crosslinking reagent sulphy-MBS (*m*-maleimidobenzoyl-*N*-hydroxysulphosuccinimide ester), followed by SDS-PAGE and fluorography. Figure 1a shows the crosslink products and their assignment to components of the import machinery as outlined below (the products containing MOM72 have slightly different apparent sizes depending on the crosslinker used; significant crosslinking to MOM30 was only observed with sulphy-MBS, and to MOM19 with DSS). When the mitochondria were pretreated with a low concentration of protease (leading to a degradation of surface proteins, but not of the membranes) only nonproductive unspecific binding of the precursor was possible^{4,7,11,14} and none of the crosslink products was generated (not shown). As MOM7, MOM8 and MOM30 are not degraded under these conditions (see below), the formation of these crosslink products must depend on the specific association of ADP/ATP carrier with mitochondria.

When isolated mitochondria and precursor were incubated at low levels of ATP and in the absence of a membrane potential, the precursor only bound to the mitochondrial surface and the insertion into the GIP site was inhibited. The exposure of this precursor on the mitochondrial surface was evidenced by its accessibility to proteases and to antibodies directed against the ADP/ATP carrier^{4,7,11-13}. The precursor was on the correct import pathway, as it followed the further import steps on addition of ATP and generation of a membrane potential^{4,7,11,12}. With this surface-accumulated precursor, only the crosslink products to MOM19 and MOM72 were formed (Fig. 1b, d). These products were identified by immunoprecipitation with specific antibodies. The apparent relative molecular masses of about 50,000 (*M_r* 50 K) and 100 K of the crosslink products fit to crosslinks between one molecule of ADP/ATP carrier (30K) and one molecule of MOM19 (19 K) or MOM72 (72 K), respectively. The crosslink product of 170 K probably consists of ADP/ATP carrier and a homodimer of MOM72, as a fraction of MOM72 is present as a noncovalent homodimer which is stable enough to survive the heating in SDS-containing sample buffer (Fig. 2; T. S. and N. P., unpublished data). As controls, none of the crosslink products was precipitated with preimmune antibodies or antibodies against porin, the most abundant outer membrane protein, whereas all were recognized by antibodies directed against ADP/ATP carrier as expected (Fig. 1b). The selective crosslinking of ADP/ATP carrier arrested at the mitochondrial surface to MOM19 and MOM72 agrees with our previous finding that both proteins are involved in the import of ADP/ATP carrier⁵ and provides strong evidence that these import receptors function by direct interaction with the precursor polypeptide. The receptor-mediated import of a chemically pure preprotein into purified mitochondria in the absence of cytosolic cofactors¹⁵ supports this conclusion.

The crosslink products with MOM7, MOM8 and MOM30 were not generated to an appreciable extent with the precursor accumulated at the receptor stage (Fig. 1b, d), suggesting that those crosslink products may be related to the GIP stage. To test this directly we made use of the observation that the precursors accumulated at the GIP site were protected against externally added protease, whereas the receptor-accumulated precursors (and the receptors themselves) were degraded by low concentrations of protease^{6,7,11,12}. ADP/ATP carrier was accumulated at the outer membrane, the mitochondria were treated with protease and then subjected to crosslinking. As expected, the crosslink products with MOM72 or MOM19 were not formed after a treatment with low protease concentrations. By contrast, the crosslink products with MOM7, MOM8 and MOM30 could still be generated (Fig. 1c, d), demonstrating that these three

† Present address: Rockefeller Research Laboratory, Sloan-Kettering Institute, New York, New York 10021, USA.

‡ To whom correspondence should be addressed.

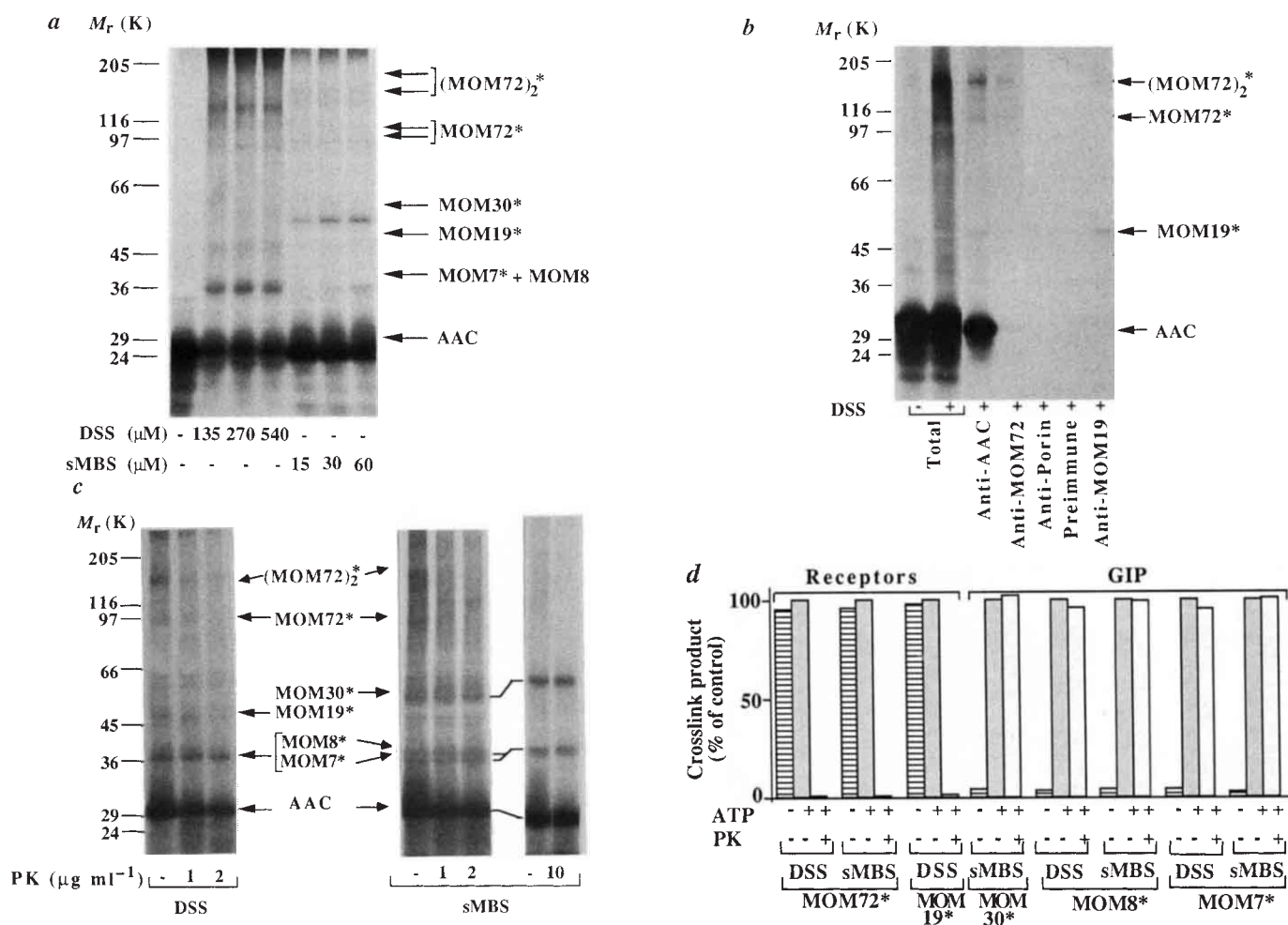


FIG. 1 Identification of mitochondrial outer membrane proteins close to the precursor of the ADP/ATP carrier. **a**, Crosslink products of the ADP/ATP carrier accumulated at the outer membrane in the absence of a membrane potential (receptor stage + GIP stage). **b**, Identification of MOM19 and MOM72 in the crosslink products on the mitochondrial surface by immunoprecipitation. **c**, Identification of MOM7, MOM8 and MOM30 in the crosslink products at the GIP stage (+PK). **d**, Stage-specific formation of crosslink products at the mitochondrial surface (−ATP) or at the GIP stage (+ATP). AAC, ADP/ATP carrier; (MOM72)₂, homodimer of MOM72; PK, proteinase K; *, products between AAC and mitochondrial outer membrane proteins.

METHODS. **a**, Rabbit reticulocyte lysate (4 μ l) containing ³⁵S-labelled precursor of AAC was incubated with isolated *N. crassa* mitochondria (25 μ g mitochondrial protein) in the absence of a membrane potential for 20 min at 25 °C as described^{4,6,10,11}. The mitochondria were reisolated by centrifugation through a 500 mM sucrose cushion and washed in SEM-buffer (250 mM sucrose, 1 mM EDTA, 10 mM MOPS, pH 7.2). Crosslinking was done in 1 mM SEM with DSS (Pierce) or sMBS (Pierce) for 30 min at 0 °C. After addition of 100 mM tris(hydroxymethyl)aminomethane (pH 7.4) or 100 mM glycine (pH 7.2) plus 0.1% (v/v) β -mercaptoethanol, respectively, and incubation for 30 min at 0 °C, the proteins were analysed by SDS-PAGE and fluorography¹¹. **b**, Isolated mitochondria and reticulocyte lysate containing precursor of AAC

were pretreated with apyrase (5 U ml⁻¹) as described^{4,11,12}, incubated for 7 min at 25 °C and further treated as described above. The samples were dissolved in buffer containing 2% SDS. The first two samples ('total') were directly subjected to SDS-PAGE, whereas the other samples were diluted 20-fold with buffer containing 1% Triton X-100 and 300 mM NaCl and immunoprecipitations with the indicated antibodies were done^{3,10,11}. Minor bands of noncrosslinked precursor were occasionally found in some precipitates independently of the type of antiserum used, apparently representing unspecific precipitation of a very minor fraction of the large amount of labelled precursor. **c**, The experiment was done as described for **a** with the following modifications. After accumulation of AAC at the outer membrane, treatment with proteinase K was as indicated^{6,11}. For the crosslinking, DSS (200 μ g ml⁻¹) and sMBS (25 μ g ml⁻¹) were used. **d**, The experiment was done as in **b** and **c** (10 μ g ml⁻¹ proteinase K). The crosslink products were quantified by laser densitometry using a calibration curve¹¹. The amount of each crosslink product obtained with AAC accumulated at receptor plus GIP stages was set to 100% (middle column of each set) (MOM72* includes the homodimer). The crosslink products with MOM7 and MOM8 were extracted (>95%) from the membranes at pH 11.5 (100 mM Na₂CO₃)^{3,4}, whereas the product containing MOM30 was carbonate-resistant (>95% in the pellet).

components are close to the GIP site. The highest efficiency of crosslinking was observed for the formation of the products with MOM7 and MOM8 (up to 15% of the accumulated ADP/ATP carrier). ADP/ATP carrier accumulated at the GIP site was extracted from the membranes at pH 11.5, suggesting that it was embedded in a hydrophilic environment⁶. The crosslink product with MOM30 was not extracted at pH 11.5, indicating that MOM30 is anchored in the outer membrane. The crosslink products with MOM7 and MOM8 were extractable at pH 11.5 (see legend to Fig. 1), suggesting that these proteins are not integral membrane proteins; they are probably associated with the outer membrane through interaction with other pro-

teins. The 38 K protein MOM38 has previously been identified as part of GIP (ref. 10). A crosslink product between ADP/ATP carrier at the GIP-stage and *N. crassa* MOM38 represented only a minor band, whereas a more efficient crosslinking was found with *Saccharomyces cerevisiae* mitochondria (T. S. and N. P., unpublished data). The considerable difference in the efficiency of crosslinking of precursors to MOM38 (=ISP42) was also observed previously^{8,16} and is obviously due to the tightly folded structure of native MOM38 (ref. 10).

Thus, at least four proteins, MOM7, MOM8, MOM30 and MOM38, seem to be close to a precursor arrested at the GIP site. On the other hand, the purified mitochondrial receptor

complex that contained ADP/ATP carrier bound to GIP consisted of the four proteins MOM19, MOM22 (a surface protein of unknown function¹⁰), MOM38 and MOM72. MOM19, MOM22 and MOM72 are protease-accessible in intact mitochondria, leaving MOM38 as the only candidate for the GIP site¹⁰. How can this apparent controversy be resolved? Because the mitochondrial receptor complex is a very labile complex that can be partially or completely dissociated under relatively mild conditions¹⁰, some components of the complex may have been lost by the purification procedure used (whereas the interaction of MOM38 with precursors arrested at the GIP stage apparently is stable enough to keep these precursors associated with the complex during the purification). When glycerol was included in the isolation buffer, the complex purified with an affinity matrix carrying anti-MOM19 antibodies contained three proteins of 7K, 8K and 30K in addition to the four known components (Fig. 2). With highly purified mitochondrial outer membranes (no detectable amounts of inner membranes), about 25 different polypeptides are resolved by SDS-PAGE^{3,4}. As expected, the three new proteins found in the receptor complex of the outer membrane represent previously unassigned outer membrane proteins (Fig. 2). Under the various conditions that

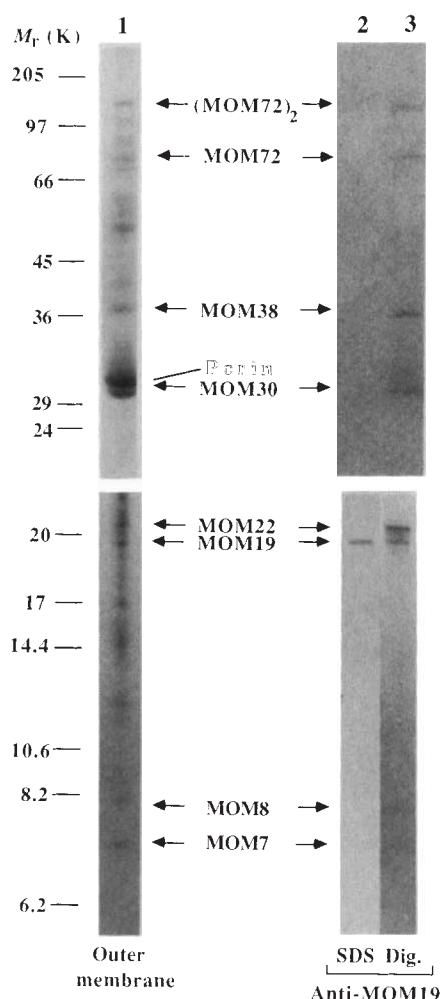


FIG. 2 Identification of three novel proteins, MOM7, MOM8 and MOM30, in the mitochondrial receptor complex. Lane 1, purified mitochondrial outer membrane^{3,4}. Lane 2, reactivity of anti-MOM19 antibodies with mitochondrial proteins under stringent conditions (denaturation in SDS followed by dilution in Triton X-100 buffer^{3,10}). Lane 3, purified receptor complex (Dig., digitonin-containing buffer).

METHODS. The purification of the mitochondrial receptor complex was done as described¹⁰ except that 10% (v/v) glycerol was included in the digitonin-containing buffer and the anti-MOM19 antibodies were covalently coupled to CNBr-activated Sepharose 4B (Pharmacia).

led to the identification of the three new components, no proteolytic degradation of proteins was observed. Proteins that are very sensitive to various proteases, in particular MOM72 and loosely folded precursor proteins, were fully stable at these conditions^{3,4,6,10}, excluding that the new components are proteolytic breakdown products.

We have thus identified three new proteins located in the outer membrane by the following methods: (1) crosslinking to ADP/ATP carrier arrested at the GIP site; (2) an improved procedure for the purification of the receptor complex of the outer membrane; and (3) comparison of the apparent sizes of the proteins with that of the components of purified outer membranes. The ADP/ATP carrier arrested at the GIP site of the outer membrane is exclusively accumulated at this stage, because a dissipation of the membrane potential completely blocks its further transport into the inner membrane and the treatment with protease degrades all precursor molecules at the receptor stage^{4,6,7,11,12}. As the purified receptor complex contains the GIP site¹⁰, the three new components identified by crosslinking are most likely to represent the three new proteins found in the receptor complex. With MOM7 and MOM8 this is also evident from their apparent size as all other outer membrane proteins resolved have apparent sizes of at least 12K. MOM7, MOM8, MOM19, MOM22, MOM30, MOM38 and MOM72 are present in the mitochondrial receptor complex in molar ratios of about 1:1:1:0.6:0.4:1:0.5 (± 0.15), respectively. Mitochondrial proteins of about 30K (28K–32K) were identified by their affinity to mitochondrial presequences^{17–20}. The relation of the 30K proteins to each other and to MOM30 is unknown.

Our study leads to the following key conclusions on the functioning of the protein import apparatus in the mitochondrial outer membrane (Fig. 3). (1) MOM19 and MOM72 act as direct receptors for the precursor polypeptides, representing the first stage of binding of precursors to mitochondria. (2) Precursor proteins are inserted into the outer membrane at the general insertion site GIP that seems to be formed by (at least) four proteins, MOM7, MOM8, MOM30 and MOM22. The relatively

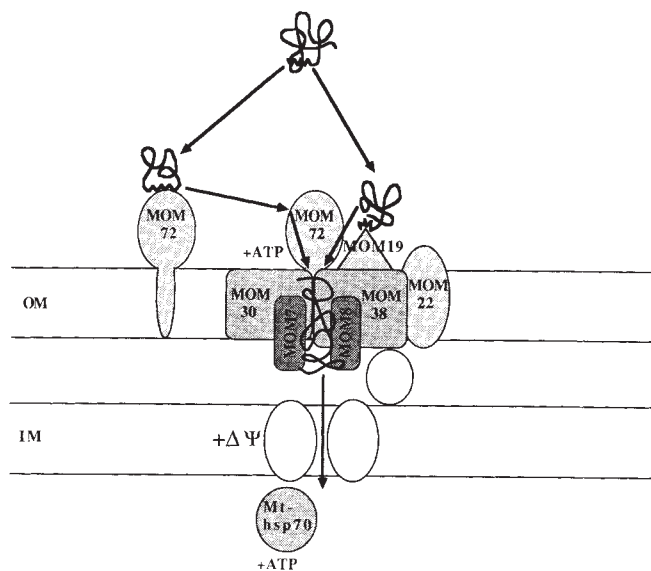


FIG. 3 Hypothetical model of the mitochondrial machinery for recognition and membrane insertion of precursor proteins. The receptor complex in the outer membrane (OM) consists of the two preprotein receptors MOM72 and MOM19, four proteins (MOM38, MOM30, MOM8 and MOM7) close to the general insertion site GIP, and MOM22 (with unknown function). The binding of preproteins to MOM72 seems to occur outside the complex, and MOM72 subsequently associates with the complex¹⁰, indicating a dynamic nature of the receptor complex. Preproteins accumulated at the GIP site are exposed to the intermembrane space²¹. Translocation into or across the inner membrane (IM) requires a membrane potential $\Delta\Psi$ (ref. 22) and the heat shock protein hsp70 in the matrix^{16,23,24}. MOMy, mitochondrial outer membrane protein with an apparent size of yK.

high efficiency of crosslinking of GIP intermediates to MOM7 and MOM8 would fit to a participation of these proteins in the formation of an insertion pore in the outer membrane. (3) The receptors and GIP are assembled in a high M_r complex of the outer membrane that consists of (at least) seven proteins. □

Received 24 June; accepted 12 November 1991.

1. Baker, K. P. & Schatz, G. *Nature* **349**, 205–208 (1991).
2. Pfanner, N., Söllner, T. & Neupert, W. *Trends Biochem. Sci.* **16**, 63–67 (1991).
3. Söllner, T., Griffiths, G., Pfaller, R., Pfanner, N. & Neupert, W. *Cell* **59**, 1061–1070 (1989).
4. Söllner, T., Pfaller, R., Griffiths, G., Pfanner, N. & Neupert, W. *Cell* **62**, 107–115 (1990).
5. Steger, H. F. *et al.* *J. Cell Biol.* **111**, 2353–2363 (1990).
6. Pfanner, N. & Neupert, W. *J. Biol. Chem.* **262**, 7528–7536 (1987).
7. Pfaller, R., Steger, H. F., Rassow, J., Pfanner, N. & Neupert, W. *J. Cell Biol.* **107**, 2483–2490 (1988).
8. Vestweber, D., Brunner, J., Baker, A. & Schatz, G. *Nature* **341**, 205–209 (1989).
9. Baker, K. P., Schaniel, A., Vestweber, D. & Schatz, G. *Nature* **348**, 605–609 (1990).
10. Kiebler, M. *et al.* *Nature* **348**, 610–616 (1990).
11. Söllner, T., Rassow, J. & Pfanner, N. *Meth. Cell Biol.* **34**, 345–358 (1991).
12. Pfanner, N., Tropsch, M. & Neupert, W. *Cell* **49**, 815–823 (1987).
13. Söllner, T., Pfanner, N. & Neupert, W. *FEBS Lett.* **229**, 25–29 (1988).
14. Pfaller, R., Pfanner, N. & Neupert, W. *J. Biol. Chem.* **264**, 34–39 (1989).
15. Becker, K., Guillard, B., Rassow, J., Söllner, T. & Pfanner, N. *J. Biol. Chem.* **267**, (in the press).
16. Scherer, P. E., Krieg, U. C., Hwang, S. T., Vestweber, D. & Schatz, G. *EMBO J.* **9**, 4315–4322 (1990).
17. Gillespie, L. L. *J. Biol. Chem.* **262**, 7939–7942 (1987).
18. Ono, H. & Tuboi, S. *J. Biochem.* **107**, 840–845 (1990).
19. Pain, D., Murakami, H. & Blobel, G. *Nature* **347**, 444–449 (1990).
20. Font, B. *et al.* *FEBS Lett.* **279**, 105–109 (1991).
21. Rassow, J. & Pfanner, N. *FEBS Lett.* (in the press).
22. Martin, J., Mahlik, K. & Pfanner, N. *J. Biol. Chem.* **266**, 18051–18057 (1991).
23. Kang, P. J. *et al.* *Nature* **348**, 137–143 (1990).
24. Ostermann, J. *et al.* *FEBS Lett.* **277**, 281–284 (1990).

ACKNOWLEDGEMENTS. We thank T. Rapoport and D. Görlich (Berlin-Buch) for discussions and U. Hanemann and A. Weinzierl for technical assistance. This study was supported by the Sonderforschungsbereich 184.

Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg–MerR

Aseem Z. Ansari*, Mark L. Chael*
& Thomas V. O'Halloran*†‡

* Department of Biochemistry, Molecular Biology and Cell Biology and

† Department of Chemistry, 2145 Sheridan Road,

Northwestern University, Evanston, Illinois 60208–3113, USA

POSITIVE control of transcription often involves stimulatory protein–protein interactions between regulatory factors and RNA polymerase¹. Critical steps in the activation process itself are seldom ascribed to protein–DNA distortions. Activator-induced DNA bending is typically assigned a role in binding-site recognition², alterations in DNA loop structures³ or optimal positioning of the activator for interaction with polymerase⁴. Here we present a transcriptional activation mechanism that does not require a signal-induced DNA bend but rather a receptor-induced untwisting of duplex DNA. The allosterically modulated transcription factor MerR is a repressor and an Hg(II)-responsive activator of bacterial mercury-resistance genes^{5–7}. *Escherichia coli* RNA polymerase binds to the MerR–promoter complex but cannot proceed to a transcriptionally active open complex until Hg(II) binds to MerR (ref. 6). Chemical nuclease studies show that the activator form, but not the repressor, induces a unique alteration of the helical structure localized at the centre of the DNA-binding site⁶. Data presented here indicate that this Hg–MerR-induced DNA distortion corresponds to a local underwinding of the spacer region of the promoter by about 33° relative to the MerR–operator complex. The magnitude and the direction of the Hg–MerR-induced change in twist angle are consistent with a positive control mechanism involving reorientation of conserved, but suboptimally phased, promoter elements and are consistent with a role for torsional stress in formation of an open complex.

The *mer* operator is located directly between the –10 and –35 hexameric RNA polymerase-binding elements of the P_T promoter. In both the repressor and activator states, MerR binds to a site centred 26 base pairs (bp) upstream from the transcription start site and in the region footprinted by *E. coli* σ^{70} RNA polymerase (RNAP)^{7,8}. This spacer DNA between the –10 and –35 hexamers of the P_T promoter is 19 bp, two bp longer and about 70° out of phase relative to a consensus *E. coli* promoter⁹. Genetic studies of deletions in the spacer region by Brown and coworkers have demonstrated the importance of a 19-bp spacer: deletion of one or two base pairs results in a stronger promoter that is no longer dependent on Hg–MerR for activation^{10,11}. Footprinting results on open complexes *in vivo* are consistent with a distortion at the centre of the operator¹².

The role of protein-induced DNA bending in this allosterically modulated DNA distortion was probed using an electrophoretic mobility shift assay^{13,14}. Analysis of the mobility of 151-bp restriction fragments with a *mer* operator at various positions (Fig. 1a) indicates that MerR induces a small bend in the operator site of $25^\circ \pm 10^\circ$ (Fig. 1b, lanes 6–10). This bend does not change on binding of the coeffector Hg(II) (Fig. 1b, lanes 11–15). Similar results were obtained with a circularly permuted 201-bp wild-type operator/promoter fragment.

Because a change in bending did not correlate with the repressor-to-activator transition, an alternative method for

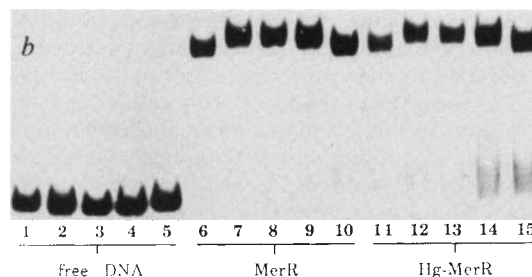
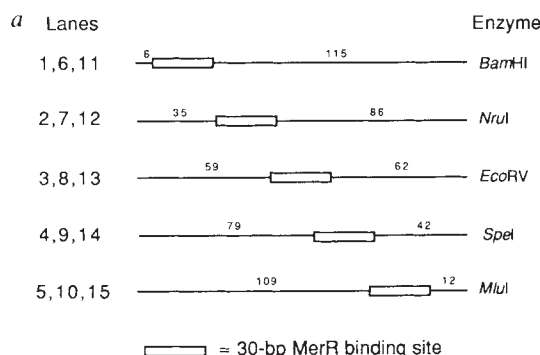


FIG. 1 Determination of MerR and Hg–MerR-induced DNA bending by circular permutation. **a**, Five 151-bp restriction fragments used in **b** indicating the relative position of the MerR-binding site in relation to the ends of the fragments. **b**, The mobility of MerR–DNA complexes using the DNA fragments listed in **a**. Gel mobility shift assays were done with 5 pM DNA (lanes 1–5), 50 nM MerR (lanes 6–10), and 0.5 μ M Hg(II) (lanes 11–15). **METHODS.** The plasmid pMB10 was made by inserting the 30-base oligonucleotide duplex: 5'-GCTTACTCCGATCATGAGTACCGAAGTAA-3' into the unique *Xba*I restriction site of the plasmid pBend2 (ref. 14) after end-filling the *Xba*I site with the Klenow fragment of DNA pol. In all, 11 151-bp restriction fragments were isolated, end-labelled with [γ -³²P]ATP and T4 polynucleotide kinase and analysed by gel mobility shift assays^{13,14}. Gel shifts were carried out in 20 μ l final volume of 10 mM Tris-HCl (pH 8), 60 mM KCl, 0.1 mM EDTA (pH 8), 5 μ g ml⁻¹ BSA, 1 mM DTT and 5% glycerol. After incubation at 23 °C for 45 min, 2 μ l 50% glycerol was added and the samples were loaded onto a 10% (75:1 acrylamide:bisacrylamide), 1 $\times$ TBE gel. Electrophoresis was for 3–5 h at 183 V. The bending locus was graphically determined using the method of Wu and Crothers¹³. The bending angle was calculated using the equation of Thompson and Landy²⁸.

‡ To whom correspondence should be addressed.

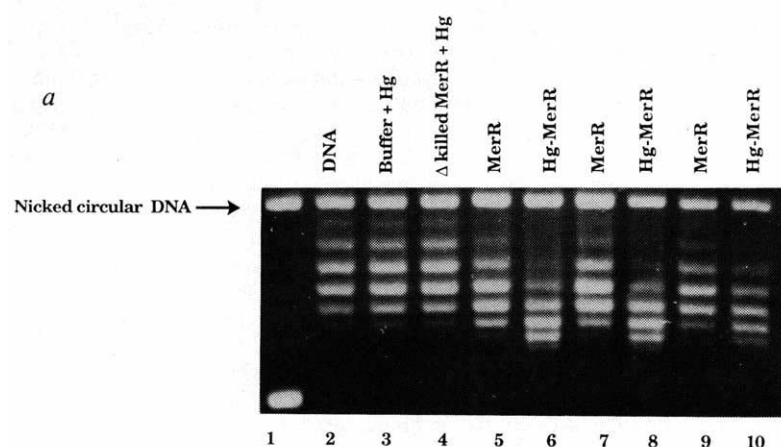


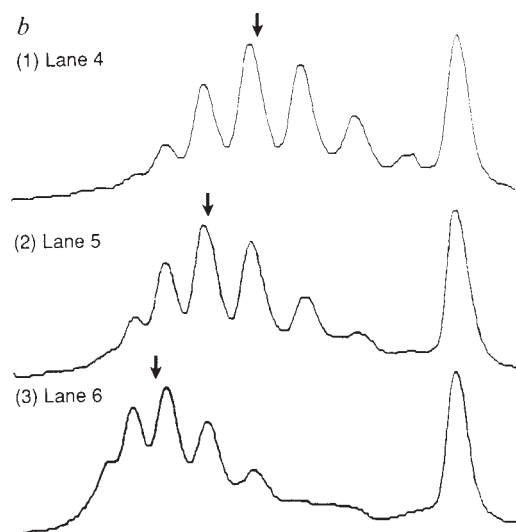
FIG. 2 Topoisomerase I assay to measure changes in helical twist mediated by MerR and Hg-MerR. *a*, Plasmid pAA15 (0.01 μ M; lane 1) was relaxed with eukaryotic topoisomerase I (lane 2) at 37 $^{\circ}$ C in the presence of: MerR storage buffer and Hg(II) (lane 3), 2.3 μ M thermally denatured Hg-MerR (lane 4), 2.3 μ M MerR (lane 5), 2.3 μ M Hg-MerR (lane 6), 0.6 μ M MerR (lane 7), 0.6 μ M Hg-MerR (lane 8), 1.2 μ M MerR (lane 9), 1.2 μ M Hg-MerR (lane 10). *b*, Densitometric traces of lanes 4, 5 and 6 from *a*. The arrow indicates the centre of the gaussian distribution. The distance between the centre of the gaussian distributions of trace 2 and trace 1 corresponds to a ΔL_k of -0.95 for MerR. Similarly, the distance between centre of distribution of trace 3 and trace 1 corresponds to a ΔL_k of -2.35 for Hg-MerR.

METHODS. DNA-MerR complexes were formed by incubating varying amounts of protein with 2 μ g plasmid in topo I buffer (20 mM Tris (pH 8.0), 1 mM MgCl₂, 1 mM DTT, 0.1 mM EDTA, 50 μ g ml⁻¹ BSA, 125 mM KCl, 5% glycerol) at 37 $^{\circ}$ C for 45 min. These complexes were incubated with topoisomerase

detecting DNA distortions was used¹⁵⁻¹⁹. A supercoiled plasmid (pAA15, plasmid construction to be published elsewhere) with 15 tandem copies of the *mer* operator, each separated by over 100 bp, was incubated with varying amounts of MerR or Hg-MerR and relaxed with eukaryotic topoisomerase I. After the removal of protein, the distribution of topoisomers was resolved by electrophoresis. Preincubation with either MerR (Fig. 2*a*, lane 5) or Hg-MerR (Fig. 2*a*, lane 6) shifts topoisomer distributions to a higher mobility relative to control samples indicating that MerR and Hg-MerR unwind the *mer* operator under the reaction conditions shown in Fig. 2. The change in the average linking number on addition of protein (ΔL_k), was determined from the shift of the gaussian centres of the topoisomer distribution (Fig. 2*b*) using the method of Depew and Wang²⁰ or Kolb and Buc¹⁵. Because writhe makes a minimal contribution to the linking number of a fully relaxed topoisomer, the ΔL_k values predominantly reflect changes in twist²¹. MerR alone yielded $\Delta L_k = -0.8 \pm 0.1$ turns or $19^{\circ} \pm 3^{\circ}$ underwinding per operator (average and standard deviation for six measurements) in the absence of coactivator. Hg(II) binding to MerR results in $\Delta L_k = -2.2 \pm 0.2$ or unwinding of the binding site by $52^{\circ} \pm 4^{\circ}$ per operator relative to control DNA. Similar unwinding is observed when pAA15 is nicked with DNase I and resealed by ligase in the presence of MerR or Hg-MerR (data not shown). The

FIG. 3 Prerelaxed DNA was used as a substrate to confirm that the topoisomerase I completely relaxes the MerR plasmid complexes. Prerelaxed pAA15 (0.01 μ M; densitometric trace 1) was further relaxed with topoisomerase I in the presence of: 1.1 μ M heat-denatured Hg-MerR (trace 2), 2.3 μ M MerR (trace 3), and 2.3 μ M Hg-MerR (trace 4). Reactions with mercury had 3 μ M Hg(II). The ΔL_k for MerR (-0.9) is obtained by subtracting trace 1 or 2 from trace 3 and ΔL_k for Hg-MerR (-2.1) was obtained by subtracting trace 2 from trace 4.

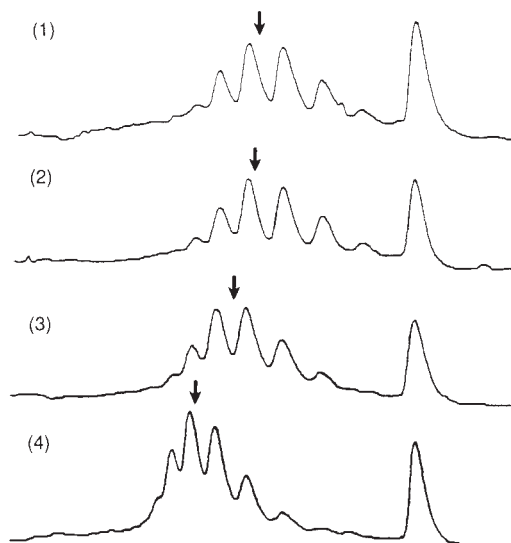
METHODS. Native supercoiled pAA15 (10.5 μ g) was treated with 10 units of eukaryotic topoisomerase I (Bethesda Research Laboratories) for 3 h at 37 $^{\circ}$ C. The reaction was terminated with a phenol-chloroform extraction, and prerelaxed pAA15 was precipitated with 3 vols ethanol, dried and then resuspended in 10 mM Tris-HCl (pH 8), 1 mM EDTA. The reaction conditions and method of analysis were identical to those described in Fig. 2.



I at 37 $^{\circ}$ C for 40 min. Calf thymus topo I (3 units; Bethesda Research Laboratories) were sufficient to relax completely the supercoiled plasmid. The reaction was stopped by the addition of SDS to a final concentration of 1.5% (phenol-chloroform extraction gave identical results). Samples were electrophoresed on 1% agarose slab gel at 4 $^{\circ}$ C in 1 $\times$ TAE with 5 mM MgCl₂ at 1.75 V cm⁻¹ for 60 h. Reactions carried out in the presence of mercury had a final concentration of 3 μ M Hg(II). Gels were stained with ethidium bromide and photographed with Polaroid type 665 film. The negatives were analysed on an LKB-GSXL laser densitometer using the Gelscan program. Several exposures were taken to ensure a linear range. Underwinding was calculated by the method developed by Depew and Wang²⁰.

observed unwinding is due to specific protein-DNA interactions as neither thermally denatured Hg-MerR (Fig. 2*a*, lane 4) nor Hg(II) in the buffer (Fig. 2*a*, lane 3) greatly alter the topoisomer distribution relative to the control plasmid. Furthermore, the vector (pSP72, Promega) without operator inserts shows no change in topoisomer distribution even at high concentrations of MerR or Hg-MerR in parallel assays. The ΔL_k values plateau after the molar ratio of dimeric protein to operator exceeds 1:1, indicating that the specific binding sites are saturated (Fig. 2). The unwinding values correspond to an equilibrium distribution of topoisomers²² under the given relaxation conditions because the same ΔL_k values are obtained when the initial template is negatively supercoiled (Fig. 2) or prerelaxed (Fig. 3).

The protein-induced bend contributes in part to the 19° unwinding angle observed for MerR alone. Because the bend angle is unchanged on addition of the allosteric effector (Hg(II)), we assume a similar contribution of bending to the 52° unwind-



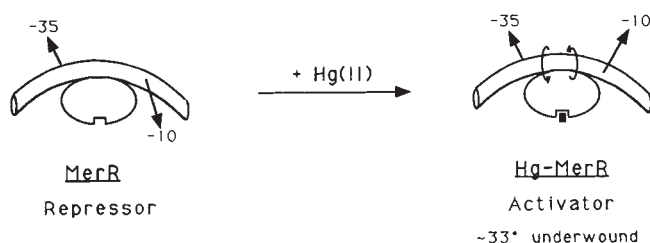


FIG. 4 Schematic diagram of the unwinding model. The operator/promoter DNA (bent cylinder) binds the MerR protein (oval) in the absence and presence of Hg(II)⁵. The arrows project from the major groove and indicate the relative positions of the -10 and -35 RNA polymerase recognition elements. The -10 element is depicted as projecting out of the page and the -35 element is projecting into the page. The arrows at the centre of the Hg-MerR-operator complex illustrate the local underwinding of the DNA helix described in the text. For purposes of illustration, the -35 element has been held constant. MerR binds one Hg(II) ion per dimer^{5,29} and mutant heterodimer experiments suggest that it binds at the interface between the two monomers³⁰.

ing angle observed for Hg-MerR. Assuming that protein-induced looping and other contributions to writhe are negligible, we attributed the 33° difference to a localized Hg-MerR-induced distortion. Enhanced reactivity of the Hg-Mer/DNA complex to intercalating chemical nucleases, such as MPE, suggest that the unwinding locus is the centre of the interpalindromic region of the *mer* operator⁶. The absence of KMnO₄ reactivity in this region indicates that the unwinding does not involve strand separation^{5,6}. This region of the promoter is an alternating purine-pyrimidine stretch and is therefore prone to helical deformations²³. Furthermore, trinucleotide analysis suggests that it is susceptible to localized bending²⁴. The free energy change of the Hg-MerR-induced DNA distortion is positive and may come at the expense of the protein-DNA binding free energy. This proposal correlates with the observed threefold decrease in the MerR-DNA binding constant in the presence of Hg(II)⁵.

The localized Hg-MerR underwinding of the spacer region coupled with the persistent bend realigns the dihedral angle between the centres of the -10 and -35 elements (Fig. 4). The resulting angle is similar to that found in a stronger promoter with an 18-bp spacer. Torsional strain in the promoter is required for isomerization from a transcriptionally inactive (closed) to the melted out (open) complex²⁵⁻²⁷. The distortion depicted in Fig. 4 is consistent with a minimal mechanism for open complex formation that involves a simple activator-mediated realignment of promoter elements as well as a possible energetic role for Hg-MerR-induced torsional stress in the DNA. □

26. Borowicz, J. A. & Gralla, J. D. *J. molec. Biol.* **184**, 587-598 (1985).
27. Ayers, D. G., Auble, D. T. & deHaseth, P. L. *J. molec. Biol.* **207**, 749-756 (1989).
28. Thompson, J. F. & Landy, A. *Nucleic Acids Res.* **16**, 9687-9705 (1988).
29. Watton, S. P. *et al. J. Am. chem. Soc.* **112**, 2824-2826 (1990).
30. Helmann, J. D., Ballard, B. T. & Walsh, C. T. *Science* **247**, 946-948 (1990).

ACKNOWLEDGEMENTS. We thank S. Adhya for pBend2, C. Aiken for suggestions and S. Fox for photographic assistance. This research was supported by the NIH.

DNA twisting and the effects of non-contacted bases on affinity of 434 operator for 434 repressor

Gerald B. Koudelka & Paul Carlson

Department of Biological Sciences, State University of New York at Buffalo, Cooke Hall, North Campus, Buffalo, New York 14260, USA

THE bacteriophage 434 repressor regulates gene expression by binding with differing affinities to the six operator sites on the phage chromosome^{1,2}. The symmetrically arrayed outer eight base pairs (four in each half-site) of these 14-base-pair operators are highly conserved but the middle four bases are divergent³. Although these four base pairs are not in contact with repressor^{4,5}, operators with A-T or T-A base pairs at these positions bind repressor more strongly than those bearing C-G or G-C⁵, suggesting that these bases are important for the repressor's ability to discriminate between operators. There is evidence that the central base pairs influence operator function by constraining the twisting and/or bending of DNA⁵⁻⁷. Here we show that there is a relationship between the intrinsic twist of an operator, as determined by the sequence of its central bases, and its affinity for repressor; an operator with a lower affinity is undertwisted relative to an operator with higher affinity. In complex with repressor, the twist of both high- and low-affinity operators is the same. These results indicate that the intrinsic twist of DNA and its twisting flexibility both affect the affinity of 434 operator for repressor.

A ring-closure assay was used to examine the twist of unbound operators with different central base sequences and to measure the extent to which wild-type and a mutant repressor can over-twist two synthetic operators (Table 1)^{8,9}. Circularizing DNA fragments containing either of these operators with T₄ DNA ligase in the absence of repressor produces two topoisomers, corresponding to linking numbers L_k of 24 and 25 (Fig. 1). The topoisomer distribution of the uncomplexed DNAs varies with the operator central sequence, the ratio of 25:24-turn topoisomer being greater for the reference than for the 7G operator (Fig. 1). Mobility shift analysis shows that these DNAs do not contain an intrinsic bend¹⁰. This difference in topoisomer distribution reveals a difference in the equilibrium twist of these two central operator sequences: the central bases of the reference operator are overtwisted relative to those in the 7G operator.

Circularizing both the reference and 7G operator-containing DNA fragments in the presence of wild-type 434 repressor increases the amount of the higher-linking-number topoisomer (Fig. 1). A mutant repressor, Phe 44 → Trp, discriminates less well between the operators than does wild-type repressor (Table 1). Circularizing the operator-containing DNAs in the presence of this mutant protein results in a much smaller increase in the amount of 25-turn topoisomer (Fig. 1). As both wild-type repressor and the Trp-44 mutant each bend DNA to identical extents (not shown), they should induce the same degree of writhe. The nonequivalent effects of these two proteins on topoisomer distribution must therefore be attributed to differences in the amount they overtwist the operator, the wild-type 434 repressor overtwisting the operator to a greater extent than the mutant.

Received 2 September; accepted 1 November 1991.

1. Ptashne, M. *Nature* **335**, 683-689 (1988).
2. Steitz, T. A. *Q. Rev. Biophys.* **23**, 205-280 (1990).
3. Lobell, R. B. & Schleif, R. F. *Science* **250**, 528-532 (1990).
4. Zinkel, S. S. & Crothers, D. M. *J. molec. Biol.* **219**, 201-215 (1991).
5. O'Halloran, T. V., Frantz, B., Shin, M. K., Ralston, D. M. & Wright, J. G. *Cell* **56**, 119-129 (1989).
6. Frantz, B. & O'Halloran, T. V. *Biochemistry* **29**, 4747-4751 (1990).
7. Brown, N. L. *et al. Molec. gen. Genet.* **202**, 143-151 (1986).
8. O'Halloran, T. V. & Walsh, C. T. *Science* **235**, 211-214 (1987).
9. Harley, C. B. & Reynolds, R. P. *Nucleic Acids Res.* **15**, 2343-2361 (1987).
10. Lund, P. A. & Brown, N. L. *Nucleic Acids Res.* **17**, 5517-5527 (1989).
11. Parkhill, J. & Brown, N. L. *Nucleic Acids Res.* **18**, 5157-5162 (1990).
12. Heltzel, A., Lee, I. W., Totis, P. A. & Summers, A. O. *Biochemistry* **29**, 9572-9584 (1990).
13. Wu, H. M. & Crothers, D. M. *Nature* **308**, 509-513 (1984).
14. Kim, J., Zwieb, C., Wu, C. & Adhya, S. *Gene* **85**, 15-23 (1989).
15. Kolb, A. & Buc, H. *Nucleic Acids Res.* **10**, 473-484 (1982).
16. Gamper, H. B. & Hearst, J. E. *Cell* **29**, 81-90 (1982).
17. Kim, R. & Kim, S.-H. *Cold Spring Harb. Symp. quant Biol.* **47**, 451-454 (1982).
18. Kim, R., Modrich, P. & Kim, S.-H. *Nucleic Acids Res.* **12**, 7285-7292 (1984).
19. Wang, J. C., Jacobsen, J. H. & Saucier, J.-M. *Nucleic Acids Res.* **4**, 1225-1241 (1977).
20. Depew, R. E. & Wang, J. C. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4275-4279 (1975).
21. Cozzarelli, N. R., Boles, T. C. & White, J. H. in *DNA Topology and its Biological Effects* (eds Cozzarelli, N. R. & Wang, J. C.) 139-184 (Cold Spring Harbor Laboratory Press, New York, 1990).
22. Amouyal, M. & Buc, H. *J. molec. Biol.* **195**, 795-808 (1987).
23. Travers, A. A. *Current Opin. struct. Biol.* **1**, 114-122 (1991).
24. Travers, A. A. & Klug, A. in *DNA Topology and its Biological Effects* (eds Cozzarelli, N. R. & Wang, J. C.) 57-106 (Cold Spring Harbor Laboratory Press, New York, 1990).
25. Buc, H. *et al. in RNA Polymerase and the Regulation of Transcription* (eds Reznikoff, W. S. *et al.*) 115-125 (Elsevier, New York, 1987).

The 25:24 topoisomer ratio of the uncomplexed 7G-containing fragment is lower than that of the uncomplexed reference operator-containing fragment, but these two DNAs form identical topoisomer distributions when complexed with wild-type repressor (Fig. 1). Repressor therefore induces a larger change in the twist of the 7G-containing fragment than it does in the reference operator-containing fragment (Table 2). By contrast, the topoisomer distributions of the two DNA fragments circularized in the presence of the Trp-44 mutant repressor vary with central operator sequence; the ratio of 25:24-turn topoisomer for the 7G operator is much smaller than that observed for the reference operator. As the free energy of twisting DNA is proportional to the square of the change in twist, the difference in affinity of the two operators for repressor must derive, in part, from the central sequence dependence of the equilibrium twist of the unbound operator.

To test this idea we used the ring closure data to calculate the ΔG of twisting the reference and 7G operators using the equation¹¹ $\Delta G = C/2 \cdot \phi^2/l$. The C term is the torsional spring constant of DNA, ϕ is the change in twist and l is the length of the DNA over which the twist is changed.

If the difference in strength of the reference and 7G operators is determined solely by sequence dependence of their equilibrium twists, the torsional rigidity of DNA, as reflected in C , is independent of central base sequence. Using this assumption, calculation of the predicted equilibrium dissociation constants indicates that the difference in affinity of the two operators for repressor should be about 12-fold less than is observed (Table 2). Repeating these calculations with larger values of C only slightly improves the agreement between the calculated and observed values (data not shown). Similar calculations based on the topoisomer distribution data obtained with Trp-44 mutant repressor also fail to account for the observed difference in affinity of the reference and 7G operators for the mutant protein (Table 2). Using different, experimentally determined, values of

the torsional spring constant for the reference and 7G operators⁸ (Table 2), the calculations based on the topoisomer data account better for the observed differences in affinity of the operators for both wild-type and mutant repressors. These results indicate that the torsional rigidity of the central base pairs is sequence-dependent.

Fujimoto and Schurr¹² have suggested that different DNA sequences have nearly identical torsional rigidities, in the light of which our results seem surprising. We therefore determined the dissociation rates of complexes formed between 434 repressor and the reference or 7G operators as a function of temperature. At 0 °C, the rate constant for dissociation (k_d) of the repressor-7G operator complex is much faster than that of the repressor-reference operator complex ($k_d^{\text{ref}} = 3 \times 10^{-3} \text{ s}^{-1}$; $k_d^{\text{7G}} = 1 \times 10^{-1} \text{ s}^{-1}$) (Fig. 2a). At 37 °C, the k_d of the 7G complex decreases 10-fold relative to that at the lower temperature, whereas that for the reference operator complex increases slightly (2.5-fold). These results show that increasing thermal energy enhances the stability of the repressor-7G operator complex, consistent with the idea that this operator contains a less flexible central sequence. Analysis of the full temperature-dependence profile of k_d (Fig. 2b) shows a favourable $\Delta H^\ddagger$ (about -2 kcal) for dissociation of repressor from the 7G operator.

In complex with the amino-terminal domain of 434 repressor, the 434 operator is overtwisted at its centre and bent towards the protein⁶. The central bases could constrain either or both of these deformations^{5,7}, but two lines of evidence indicate that they influence DNA twist. First, an earlier investigation showed that two operators with different central sequences do not contain an intrinsic bend¹⁰, whereas here we show that these operators do have different equilibrium twists. Second, the mutant repressor, which does not discriminate so well between base changes at the centre of the operator, also overtwists the operator less but bends it to the same extent as wild type.

Base composition effects on DNA twist could influence operator affinity for repressor by (1) limiting the degree that the operator can be twisted in the protein-DNA complex; (2) influencing equilibrium twist of the unbound operator; or (3) altering the ease with which the operator DNA can be deformed into the properly overtwisted configuration for complex formation (altering torsional flexibility). Our data do not support the first hypothesis. The data presented here indicate that both the

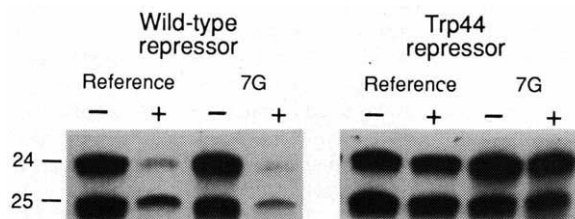


FIG. 1 Topoisomer distribution of circularized reference operator- and 7G operator-containing DNA fragments. Labelled linear DNAs were circularized in the absence or presence of wild-type 434 repressor or 434 repressor bearing a Phe 44 $\rightarrow$ Trp substitution. For reactions in the presence of binding protein, each was present at 1 μM . A portion of the autoradiograph of a representative gel is shown. The positions of bands corresponding to topoisomers of L_k values of 24 and 25 are indicated. Determination of linking number has been described⁸.

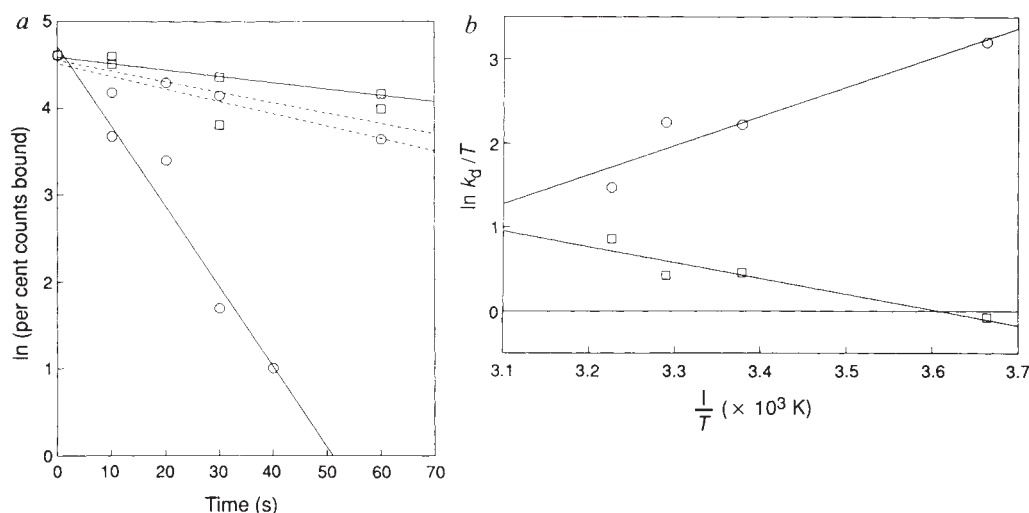
METHODS. Dephosphorylated 258-base-pair linear DNA fragments containing a single 434 operator and *Ban*I cohesive ends were isolated by gel from plasmids constructed as in ref. 10. The 5' end was labelled by incubating DNA with polynucleotide kinase and [γ -³²P]ATP, followed by a chase with cold ATP. Fragments were purified away from nucleotides and kinase by extraction with phenol followed by precipitation with ethanol. Labelled DNA fragments were incubated in 10 mM Tris-HCl, pH 7.8, 50 mM KCl, 1 mM MgCl₂, 10 mM DTT, 50 $\mu\text{g ml}^{-1}$ bovine serum albumin at 0 °C for 10 min in the absence or presence of binding protein. Sufficient DNA ligase was added to this mixture to give > 50% ligated circles after a further 10-min incubation. Reactions were quenched by extraction with phenol followed by precipitation with ethanol, or by adding a stop solution containing sufficient SDS and EDTA to give a final concentration of 0.5% and 50 mM, respectively. Reaction products were fractionated by electrophoresis at 18 V cm^{-1} in 5% polyacrylamide gels (29:1 monomer:methylene bisacrylamide) containing 90 mM Tris-HCl, 90 mM sodium borate and 1 mM EDTA. Autoradiography was by exposure of the dried gel to preflashed Kodak XAR-5 film with an intensifying screen at -70 °C. Reaction products (topoisomers, nicked circles and linear dimers) were identified as described⁹.

TABLE 1 Affinities of wild-type and Trp-44 mutant 434 repressors for synthetic 434 operators

Operator	Sequence	Relative K_d	
		Wild type	Trp-44
Reference	ACAATATATATTGT	1	1
7G	ACAATAGCTATTGT	60	4

Binding affinities (K_d) are expressed as the concentration of repressor monomers needed to occupy half-maximally each operator in a DNase I protection experiment. These values have been normalized to the amount of repressor needed to half-maximally occupy the reference operator, which was determined in a parallel experiment. Under the conditions used in these experiments, unity for wild-type repressor was $2 \times 10^{-8} \text{ M}$, and for Trp-44 repressor, $3 \times 10^{-7} \text{ M}$. Synthetic 434 operator-containing plasmids were constructed as described⁹. An operator containing the *Eco*RI-*Hind*III fragment, 3' end-labelled at the *Eco*RI end, was purified by gel electrophoresis. DNase I protection experiments were done essentially as described¹⁹, using assay conditions of 10 mM Tris, pH 7.8, 50 mM KCl, 1 mM MgCl₂, 2.5 $\mu\text{g ml}^{-1}$ sonicated chick-blood DNA, 100 $\mu\text{g ml}^{-1}$ bovine serum albumin, 1 nM operator DNA and repressor at concentrations that varied by increments of 1.5-fold for each titration. Following incubation at 25 °C for 15 min sufficient DNase I was added to give, on average, one cleavage per DNA molecule in 5 min of further incubation. Autoradiographs were quantified by densitometry. Site-directed mutagenesis of 434 repressor was done as described²⁰. The mutant protein bearing the Phe 44 $\rightarrow$ Trp substitution was purified as in ref. 21.

FIG. 2 Temperature dependence of the rate of dissociation (k_d) of 434 repressor from reference- and 7G operator-containing DNA. *a*, The log of the per cent of total counts bound at different times after addition of 1,000-fold excess of cold operator-containing DNA to a reaction mixture containing labelled operator containing DNA (~1 nM) and repressor at a concentration sufficient to half-maximally bind the labelled input DNA to a nitrocellulose filter. Points plotted are averages of three experiments. Solid lines are data obtained at 0 °C, dashed lines represent data at 37 °C. □, Reference operator; ○, 7G operator. *b*, Arrhenius analysis of the k_d data: $\ln k_d/T$ is plotted as a function of $1/T$. □, Reference operator; ○, 7G operator.



equilibrium twist and torsional rigidity of DNA are sequence-dependent. Calculations based on the topoisomer distribution data show that the difference in total twist induced by wild-type 434 repressor on the two operators cannot completely account for central base sequence effects on operator strength. If different base compositions have slight, but significantly dissimilar, torsional rigidities, this would allow topoisomer data to predict the observed 60-fold difference in operator strength.

Consistent with the idea that the central bases could be affecting the ease with which the operator can be overtwisted, a correlation has been¹³ established between sequence-depen-

dent flexure and operator affinity for 434 repressor, although twist and bend effects could not be distinguished. The suggestion that sequence-dependent torsional rigidity of DNA influences operator affinity for repressor is in qualitative and quantitative agreement with the proposals of Hogan and Austin¹⁴. These analyses and our data all agree with results showing that 434 repressor binds better to a more flexible, nicked operator than to an intact operator⁷.

The suggestion that the torsional rigidity of DNA varies with base sequence seems to be at odds with the proposals of Fujimoto and Schurr¹². Further inspection of their data shows, however, that the difference in average upper and lower bound torsion constants for different DNA sequences is as large as 2.6-fold. This is greater than the two-fold difference in C of the reference and 7G operators which was used to calculate the ΔG that accounts for the observed binding data.

In addition to 434 repressor, the sequence dependence of DNA structure has been proposed to affect the structure of complexes of 434 Cro, CRP, *trp* repressor and *Escherichia coli* RNA polymerase with their specific DNA sites¹⁵⁻¹⁸. Our results highlight the importance of considering the effects of base composition not only on equilibrium structure of DNA and protein-DNA complexes, but also on the sequence-dependent dynamics of the DNA molecule. □

TABLE 2 Energetics of DNA twisting as a determinant of operator affinity for 434 repressor

	Wild-type repressor		Trp-44 repressor	
	Reference	7G	Reference	7G
ΔT_w (deg)	6.64	14.19	1.58	1.72
C^*	2.9×10^{-19}	2.9×10^{-19}	2.9×10^{-19}	2.9×10^{-19}
ΔG (kcal)	0.21	1.17	0.012	0.014
k_d (7G)/ k_d^{ref}	5.1	5.1	1.00	1.00
C^*	3.3×10^{-19}	6.7×10^{-19}	3.3×10^{-19}	6.7×10^{-19}
ΔG (kcal)	0.23	2.69	0.013	0.032
k_d (7G)/ k_d^{ref}	64.1	64.1	1.03	1.03

The variables required to complete this calculation were determined as follows. The helical twist of the entire fragment was determined from the topoisomer ratios and the equations in refs 8 and 9. The relative amount of each topoisomer was determined from counting the excised bands and/or densitometric quantitation of the autoradiograph. The values used are the average ratios of the 25:24-turn topoisomers determined in 5 experiments. The difference in twist of the central sequence between the unbound and bound operators (ΔT_w) was determined from the helical twist of the entire fragment using a weighted average. The table depicts the results of calculations that assume the repressor-induced twist is distributed over the central four bases in the operator. For the results given in the upper half of the table, the value of the torsional spring constant is assumed to be 2.9×10^{-19} erg cm (see text). This value has been reported before^{8,9} and may be close to an upper limit for mixed-sequence DNA²². The lower half of the table summarizes the results of calculations using different, experimentally determined, values of C for the reference and 7G operators. These values were deduced from the temperature dependence of the topoisomer distribution of the two circularized DNA fragments. For all calculations, two additional assumptions were made: (1) the repressor-induced twist is distributed over the four central base pairs, and (2) the number of bases overtwisted by repressor is independent of central base composition. These assumptions are supported by X-ray crystallographic data and binding measurements indicating that the noncontacted region of the operator is 4 base pairs long⁵⁻⁷.

* Value of C is given in erg cm.

Received 23 May; accepted 17 October 1991.

- Wharton, R. P., Brown, E. L. & Ptashne, M. *Cell* **38**, 361-369 (1983).
- Ptashne, M. *A Genetic Switch* (Blackwell, Palo Alto, 1986).
- Wharton, R. P. thesis, Harvard Univ. (1985).
- Anderson, J. E., Harrison, S. C. & Ptashne, M. *Nature* **326**, 888-891 (1987).
- Koudelka, G. B., Harrison, S. C. & Ptashne, M. *Nature* **326**, 886-888 (1987).
- Aggarwal, A., Rodgers, D. W., Drottler, M., Ptashne, M. & Harrison, S. C. *Science* **242**, 899-907 (1988).
- Koudelka, G. B., Harbury, P., Harrison, S. C. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4633-4637 (1988).
- Shore, D. & Baldwin, R. L. *J. molec. Biol.* **170**, 983-1007 (1983).
- Horowitz, D. & Wang, J. C. *J. molec. Biol.* **173**, 75-91 (1984).
- Koudelka, G. B. *Nucleic Acids Res.* **19**, 4115-4119 (1991).
- Barkley, M. D. & Zimm, B. H. *J. chem. Phys.* **70**, 2991-3007 (1979).
- Fujimoto, B. S. & Schurr, J. M. *Nature* **344**, 175-178 (1990).
- Drew, H. R., McCall, M. J. & Calladine, C. R. in *Biological Aspects of DNA Topology* (eds Cozzarelli, N. & Wang, J. C.) 1-56 (Cold Spring Harbor Laboratory, New York, 1990).
- Hogan, M. & Austin, R. *Nature* **329**, 263-266 (1987).
- Mondragon, A. & Harrison, S. C. *J. molec. Biol.* **219**, 321-334 (1991).
- Dalma-Weiszhausz, D. D., Gartenberg, M. R. & Crothers, D. M. *Nucleic Acids Res.* **19**, 611-616 (1991).
- Otwinowski, Z. *et al. Nature* **335**, 321-329 (1988).
- Auble, D. T. & deHaseth, P. L. *J. molec. Biol.* **202**, 471-482 (1988).
- Johnson, A. D., Meyer, B. J. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5061-5065 (1979).
- Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488-492 (1985).
- Anderson, J. A., Ptashne, M. & Harrison, S. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1307-1311 (1984).
- Taylor, W. H. & Hagerman, P. *J. molec. Biol.* **212**, 363-376 (1990).

ACKNOWLEDGEMENTS. We thank P. Hagerman for assistance in analysing the topoisomer distribution data and S. Harrison, D. Rodgers and L. Shimon for data in advance of publication and for discussion. This work was supported by a grant from the NIH to G.B.K.

What's new in '92

New products for the new year include a system for rapid separation and membrane transfer without blotting, a range of cell culture flasks for the selection of specific cell phenotypes and a PC-based nuclear structure physics software program.

NOVAGEN'S Ig-Prime System is designed for rapid **cloning of immunoglobulin (Ig) variable region cDNAs from human, mouse and rabbit sources** (*Reader Service No. 100*). The system, Novagen says, takes advantage of the fact that mammalian light and heavy chain Igs contain conserved regions adjacent to the hyper-variable complementarity defining regions (CDRs). Specially designed primers are used to amplify the CDRs of light and heavy chain Igs, typically starting with less than 2 µg total RNA. The amplification products are cloned by ligation into the pT7Blue T-vector and transformation into NovaBlue competent cells. The cloned inserts are sequenced by using the primers supplied and following standard methods. The method allows for the rapid analysis of hybridomas, myeloma cell progression and heterogeneity, idiotype networks, B-cell clonal ontogeny, and phylogenetic relationships. In addition, the characterized Ig-Prime V_H and V_L products can be used in the subsequent construction of recombinant/chimaeric antibodies. Each kit contains sufficient components for 20 sets of amplifications, ligations and transformations.

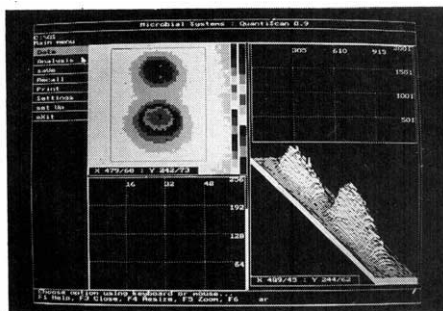
Serva is distributing a new line of Biotek Parcosil/Parcomer **HPLC columns**, which are designed specifically for the separation of proteins and peptides in both analytical and preparative procedures (*Reader Service No. 101*). The packing materials available are based on spherical silica (Parcosil) and on full polymer (Parcomer). The polymer is a polymethacrylate. Parcomer/Parcosil phase materials are available for reverse phase (C4, C8, C18), ion exchange (strong and weak, cationic and anionic) and for hydrophobic interaction chromatography. Except for the reverse-phase materials, a special surface chemistry is applied to the particles, which minimizes unwanted unspecific interaction of the probe molecules with the surface of the phase material, Serva says. This feature is designed to provide high recovery rates and good resolution. Primary applications include the separation of native and synthetic peptides, a wide range of proteins, native and synthetic nucleic acids, and smaller (pharmaceutical) molecules. As the polymer-based materials are stable over a wide pH range (pH 1–12), "cleaning-in-place" procedures using

NaOH are possible. The Parcosil/Parcomer columns are supplied in various pore and particle sizes to suit many HPLC and FPLC applications. Pre-packed columns (both steel and PEEK biocompatible polymer) are available in a range of column dimensions.

Densitometry

Hoefer Scientific is the new UK distributor for the Elscript range of high-resolution **densitometers** made by the German company, Hirschmann (*Reader Service No. 102*). Three models are available, with optical systems that are suitable for both one- and two-dimensional scanning. The instruments offer the user a choice of transmission only, transmission and reflectance or UV and fluorescence. Optimized resolution and sensitivity for a variety of gel and blot staining methods is provided by the wide variety of slit sizes and filters. IBM-compatible software for control of the instruments and for data manipulation is also provided. Results and graphical output can be exported to TIFF files for incorporation into wordprocessing and desktop publishing documents.

According to Biosoft, the company's QuantiScan **densitometry software** transforms an IBM PC or compatible computer and a desktop scanner into a densitometer at an affordable price (*Reader Service No. 103*). QuantiScan



Densitometry for the budget conscious.

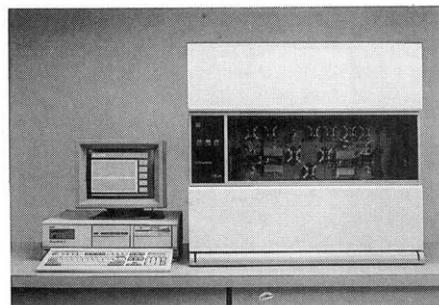
accepts 4-bit (16 greyscale) and 8-bit (256 greyscale) Intel format TIFF files produced by most makes of desktop scanner or video camera. Densitometry peaks can be located in manual, tangential and automatic modes. Integration of peaks can be performed in the X or Y direction and the results displayed in tabular format. Two-dimensional (2-D) integration can be performed enabling location of 2-D peaks. In addition, smoothing, edit-

ing and background subtraction can be carried out on densitometer plots. Results can be either exported or stored as ASCII files. Screen information can also be sent to printers or exported in Paint and Paintbrush formats. QuantiScan, which runs on an IBM PC/XT/AT/PS2 or compatible computer, costs £249 (UK) or \$499 (US). Demonstration disks are available on request.

Ideas in instrumentation

The new **3-D cytometer workstation** from Biological Detection Systems is designed to use the latest techniques in removing out-of-focus information (also known as digital confocal microscopy) inherent in many biological and non-biological samples (*Reader Service No. 104*). Coupled with this package is the ability for researchers to take their deconvoluted data sets and render them in three-dimensional space. These data sets can be animated to provide 3-D movie sequences. The 3-D Cytometer accepts input from conventional optical confocal microscopes for complete 3-D rendering. The entire system resides inside a standard Macintosh computer and can run with or without hardware acceleration. This system can be used as a stand-alone 3-D package or as part of the BDS family of research cytometers.

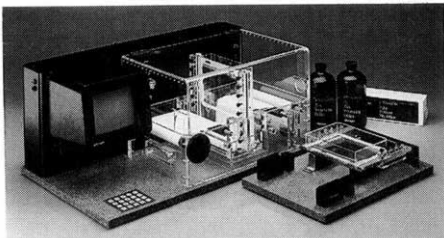
With the new ProSequencer advanced micro **protein sequencing system** from Millipore users can sequence both covalently attached and adsorbed samples, completely in the liquid phase, using the same instrument (*Reader Service No. 105*). Sequencing performed by an experienced user is facilitated by the single point of control for all modules (Edman unit, HPLC and data analysis). A positive-displacement reagent delivery system, reduced system volume and



Millipore's ProSequencer for sequencing in covalent and adsorptive modes.

reduced reagent consumption, combined with a microbore PTH-amino acid analysis system provide high sensitivity and sub-picomole sequence analysis in both adsorptive and covalent modes, Millipore says. The ability to sequence in both modes allows the user to adapt new chemistries as they become available. In addition, patented covalent attachment chemistries provide the user with numerous possibilities to perform pre- and post-analysis chemistries.

The new AutoTrans system from Betagen combines **electrophoretic separation of macromolecules and membrane transfer** into one integrated process (*Reader Service No. 106*). Nucleic acid or



Two processes, one instrument.

protein samples are loaded on a short vertical acrylamide or agarose gel. As the molecules separate and elute from the bottom of the gel, the patented moving web system advances the transfer membrane across the gel surface and binds the eluted molecules. Two models are available. The AutoTrans 700 is a self-contained system that comprises a computer control and power module in addition to the electrophoresis and roller assembly. The microprocessor electronics control and monitor all systems operations, including the integral multi-range power supply and field-inversion controller. The computer also provides essential timing, speed and coordination of the web-drive motor and buffer recirculating pump. A 24-line CRT provides an easily readable interface to the system with commands entered through the sealed keypad. The model 700 has a variety of uses, such as dideoxy or multiplex DNA sequencing, genetic or physical mapping studies, or preparation of Southern or western blots. In some applications, multiple sample sets can be loaded onto the same gel to provide continuous, rapid sample processing, Betagen says. AutoTrans 350, a non-computerized version consisting of an electrophoresis and roller module plus a web-drive controller, is designed for repetitive, single applications. The 350 model, however, can be upgraded should user requirements change.

The UV Band-Eluter E91 from Biometra is designed to simplify the **elution of DNA/RNA fragments from preparative gels** (*Reader Service No. 107*). Combining

a UV-transparent horizontal electrophoresis unit with an integral UV-light source allows the separation of nucleic acids to be monitored directly in the presence of ethidium bromide. A special geometry of sample and trap slots ensures the migration of the sample directly into the trap slots. A focusing solution, which is pipetted into the traps, provides a sharp focusing of bands in front of the trap, Biometra says. The desired fragment moves into the trap as a sharp band and can easily be removed with a pipette. Recovery is controlled online: if the fluorescence is totally removed from the trap, the recovery is close to 100 per cent. By repeating the steps, a complex mixture of nucleic acids can be fractionated band by band. Aside from its preparative function, the UV Band-Eluter is a useful tool for analytical agarose gel electrophoresis of nucleic acids. The apparatus offers several advantages: handling of ethidium bromide is restricted to the apparatus, contamination of darkrooms and UV tables is minimized, separation can be monitored directly during the run, bad runs due to incorrect run parameters can be altered immediately, and enzymatic reactions can be monitored directly during the reaction.

Kits and reagents

The cDNA ClonStruct Kit from United States Biochemical is designed to promote efficient **cloning of full-length cDNAs** (*Reader Service No. 108*). The kit comes complete with a vector-primer, in addition to all the necessary enzymes and reagents to perform five cDNA library constructions. A single library construction requires 10–20 µg of poly A⁺ RNA, USB says. The procedure is as follows: a modified vector serves as a primer for first strand synthesis; each end of the vector-cDNA:mRNA hybrid is modified by the addition of a C tail using terminal deoxynucleotidyl transferase (TdT) — a reaction that has been calibrated to add a tail of 5–20 residues; the vectors contain a *Bst*X I restriction site whose recognition sequence internally contains six G residues. *Bst*X I restriction removes the C tail from the non-hybrid end of the molecule while exposing a tail of four G residues; the vector containing the cDNA is recircularized using T4 DNA ligase; a second strand synthesis and repair reaction is mediated by the action of DNA polymerase I, ribonuclease H and T4 DNA ligase; and the cDNA clones are transformed into an *Escherichia coli* host to generate the library. USB outlines several advantages to this strategy: it generates a large-size library where the cDNAs are inserted in a known orientation into a vector as they are synthesized; ligation of the vector-cDNA:mRNA occurs efficiently by an intramolecular reaction; and TdT, used to tail the vector-

cDNA:mRNA molecule, acts more efficiently on a blunt end than on a 3' recessed end. There are two sets of modified vectors to choose from: pXPRS +/- (eukaryotic cell expression vectors) and pTRXN +/- (for *in vitro* transcription using T7 RNA polymerase). Other vector-primers may be purchased separately for use with the \$495 (US) cDNA ClonStruct Kit.

Agarose gels to suit a variety of applications are now available as part of the SepRate range of **agaroses** from Amersham (*Reader Service No. 109*). SepRate-DNA is a standard agarose for the separation of DNA fragments. Each batch is tested in a Southern blot with Hybond-N membrane to ensure compatibility. For the separation of small DNA fragments of less than 1,000 base pairs, Amersham recommends its SepRate-SDF agarose. High gel strength and melting temperature facilitate handling, Amersham says. SepRate-PFGE, an agarose formulated for pulsed-field gel electrophoresis, is designed to enhance the speed of separation of fragments in the range of 50 kilobases to 6 megabases. Amersham claims that SepRate-PFGE can save up to 50 per cent of running time without loss of resolution. The agarose is DNase- and RNase-free. Completing the line up is SepRate-LMP, a low melting



Applications-tested agaroses.

point agarose designed for the easy recovery of DNA fragments. As the melting point of SepRate-LMP agarose is below 65 °C, both resolved proteins and nucleic acids can be recovered without denaturing, Amersham says. All four agaroses are available in a range of pack sizes to suit requirements, including a special 2 g trial size.

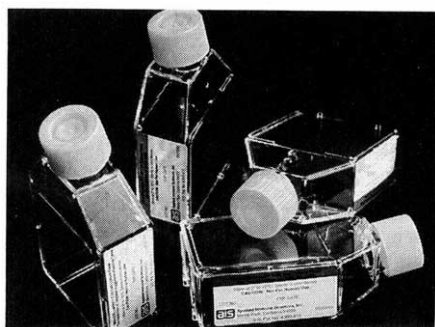
For the most up-to-date information on Bachem's line of **synthetic peptides, substrates, protected amino acids and resins**, pick up a free copy of the company's new 1991/1992 catalogue (*Reader Service No. 110*). New product listings include highly specific V₁ and V₂ vasopressin antagonists, highly potent tachykinin antagonists, a wide selection of beta-amyloid peptide fragments, scorpion neurotoxins and new immunoadjuvants. In addition to product listings, the catalogue includes a wealth of information on special services provided by Bachem, such as custom synthesis, the preparation of GMP-grade products with supporting documentation

for clinical applications, and bulk quantity pricing.

Triacsin C, a **potent inhibitor of long-chain acyl-CoA synthetase** and a pivotal enzyme in the regulation of metabolic pathways and the biosynthesis of complex lipids and acylated proteins, is now available from Kamiya Biomedical (*Reader Service No. 111*). The enzyme is expected to play an important role in studies on growth regulation, lipid metabolism, protein biosynthesis, biosynthesis of complex lipids, regulation of metabolic pathways, and in intracellular protein transport.

Cell science

The new AIS MicroCELLector cell culture flasks from TechGen allow specific cell phenotypes to be positively selected or depleted through specific monoclonal antibodies bound to the surface of the flasks (*Reader Service No. 112*). Using proprietary attachment technology, specific cell receptors are bound to the surface of the cell culture flask. When samples of blood or bone marrow are added, the immobilized receptors bind to the surface antigens of the targeted cells; cells not bearing the recognised antigens remain in suspension. Separation should take place in less than 60 minutes. According to TechGen, the MicroCELLector flasks offer a specificity that approaches that of flow cytometry but with the convenience and cost effectiveness of a disposable



Culture flasks for specific cell binding.

system. Alternatively, the flasks can be used to complement flow cytometry as a pre-enrichment step before cell sorting. There are specific MicroCELLector flasks for T, B and progenitor (stem) cell preparation. In addition, an "activated" flask is available that enables researchers to bind their own antibody. As an alternative to the activated flask, TechGen offers streptavidin and goat anti-mouse flasks.

Researchers studying chromosomal abnormalities may be interested to hear that Gibco BRL has introduced AmnioMAX-C100, a new medium designed specifically for the *in vitro* propagation of primary cultures of human amniotic fluid cells (*Reader Service No. 113*). AmnioMAX-C100, which is a basal medium that has only one pre-measured supplement added before use, provides maximum growth rates of amniocytes and an increased metaphase yield, the company says. Each lot is qualified for use on primary amniotic fluid isolates using a six-day colony assay by a commercial cytogenetic reference laboratory.

Dealing with data

A new image analysis software program for cytogenetic quantification has been introduced by Seescan to help cytogeneticists in the preparation of karyograms (*Reader Service No. 114*). Running on Seescan's image analysis system, the software features a range of functions, including morphological and densitometric measurements, texture analysis, chromosome counting and karyotyping, and ideograms. Designed to replace cut-and-paste techniques, the karyotyping program works as follows: having imaged a metaphase, the user points to the centromere of each chromosome and cuts between touching and overlapping chromosomes using the mouse. The system then generates a karyotype, sorting chromosomes on the basis of length and arm ratio and lining up the centromeres. At this point, the user can rotate chromosomes, delete and add chromosomes, and change their position within the karyotype. In addition, text and a scale bar may be added, and the resultant image stored on disk, exported to desktop

publishing packages, or printed. The morphology program can be used to make a wide range of measurements (area, perimeter, length and breadth) on cells, nuclei and chromosomes. Using the densitometry program, optical density measurements of cell nuclei can be made to assess the DNA content of stoichiometrically stained material.

Version 3.0 of the FCC Nuclear Structure Software program is now available from TransTech (*Reader Service No. 115*). The software is both a general introduction to nuclear structure physics and a useful research tool for exploring the three-dimensional (3-D) geometry of the atomic nucleus. The program allows the user to construct nuclei in 3-D space — either using default structures by specifying only the number of protons or neutrons, or by building nuclei by hand, nucleon by nucleon. The position of valence nucleons can then be altered to test the spin, parity, binding energy, radius, Coulomb energy, and magnetic and quadrupole moments of various configurations of protons and neutrons. Numerical comparisons are made among the shell, liquid drop and FCC models of nuclear structure and displayed together with the graphical representation. The software runs on all IBM (PC/XT/AT/PS2) computers and compatibles with EGA or VGA colour graphics capabilities. A hard disk and mouse are recommended, but are not required. The \$290 (US) FCC Nuclear Structure Software program is available on either a 3.5- or 5.25-inch disk.

SWEEP, VACCINE and D-FENCE are three anti-virus products from the Sophos armoury (*Reader Service No. 116*). SWEEP is a virus-specific software package that scans disks and networks for all "known" viruses, alerting the user if any are detected. The program is updated monthly to provide users with the latest virus patterns. SWEEP has a built-in removal facility that allows the user to delete "infected" files or immobilize infected sectors. VACCINE, however, is virus-non-specific. Any change, however caused, will be detected. Completing the line up is D-FENCE, a memory-resident utility that prevents the use of unauthorized floppy disks, thereby reducing the chance of virus infections from the outside. Unlike conventional access control packages, D-FENCE does not interfere with normal operation of the PC, Sophos says. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

ADVERTISEMENTS

Your Valuable Protein

MICHAEL WEST CONSULTING

Toronto, Ont.

provides top Quality Service:

- PURIFICATION • PROTEOLYSIS
- PEPTIDE SEPARATION
- SEQUENCING

CONSULTATION
AND TRAINING

..... ON SITE!

PHONE

FAX

416-255-9288

416-979-5294

NH₂-Ala-Glu-Asn-Ser.....
YOUR SEQUENCE

Reader Service No. 13

Synthetic Oligonucleotides

Custom made probes and primers,
synthesised, purified and shipped in 3 days

£3.50 per base

(All inclusive)

0.2 µM scale (approx. 500 µg)

Modified bases, biotin or digoxigenin labelling,
data base analysis and full design facilities

Molecular Medicine Unit,
King's College School of Medicine & Dentistry,
Denmark Hill, London SE5 8RX

Tel. XX-44-(0)71 - 326 3126

FAX XX-44-(0)71 - 733 3877

Reader Service No. 56